AD_____

Award Number: DAMD17-00-1-0025

TITLE: Cladistic Association Analysis of Genetic Effects on Prostate Cancer in African Americans

PRINCIPAL INVESTIGATOR: Rick A. Kittles, Ph.D.

CONTRACTING ORGANIZATION: Howard University Cancer Center
Washington, DC 20060

REPORT DATE: March 2002

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	March 2002	Final (15 Feb 00 - 14 Feb 02)

4. TITLE AND SUBTITLE
Cladistic Association Analysis of Genetic Effects on Prostate Cancer in African Americans

5. FUNDING NUMBERS
DAMD17-00-1-0025

6. AUTHOR(S)
Rick A. Kittles, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Howard University Cancer Center
Washington, DC 20060

rkittles@howard.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. Abstract ***(Maximum 200 Words)*** *(abstract should contain no proprietary or confidential information)*
The goals of this project are to (1) provide a formal evaluation of variation at three candidate gene regions for prostate cancer (the androgen receptor gene, 5α-reductase type 2 gene, and the chromosome X region Xq27-28) in 2,000 clinically evaluated unrelated men, and (2) exploit the evolutionary history of haplotypes in order to determine if haplotypic differences account for phenotypic variation in prostate cancer, prostate specific antigen (PSA) levels and disease progression. To date we have collected blood and clinical data from about 1700 unrelated men with and without prostate cancer. Molecular genotyping of three polymorphisms within the androgen receptor has been performed in 1700 individuals and three polymorphism within the 5α-reductase type 2 gene in about 1100 of the subjects. In addition, five microsatellites have been typed within the candidate gene region in and around Xq27-28 for all subjects. Preliminary analyses reveal significant linkage disequilibrium between the androgen receptor markers in African Americans and suggest a strong correlation between specific androgen receptor haplotypes and prostate cancer. The other two gene regions have yet to be analysed for association with prostate cancer.

14. SUBJECT TERMS
genetic susceptibility, genetic polymorphic markers, haplotypes, androgen receptor gene, chromosome Xq27-28 region, 5α-reductase type 2 gene

15. NUMBER OF PAGES
33

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Table of Contents

INTRODUCTION

For unknown reasons, prostate cancer incidence and mortality rates for African American males are among the highest in the world. Very few hereditary prostate cancer studies have included African Americans. This is unfortunate; especially since the population history Africans Americans is quite different than other populations. Thus, genetic predisposition to a common disease like prostate cancer may also be different. The identification of susceptibility genes will provide insight into critical rate limiting steps in the carcinogenic pathway of both inherited and sporadic cases of the disease. The specific goals of this project are as follows: (1) Extraction of genomic DNA from blood collected from 2000 unrelated men (1500 African Americans and 500 European Americans) from Columbia, South Carolina; Chicago, Illinois; and Washington, D.C. (2) The genotyping of microsatellite (STRs) loci and single nucleotide polymorphisms (SNPs) in order to construct compound haplotypes from three candidate genomic regions. (3) Analyze the effects of differences between haplotypes on the vulnerability to prostate cancer and related PSA levels using cladistic association analysis (Templeton et al., 1987). Our expectations for this project are to determine if any of the candidate gene regions from a large sample of clinically evaluated and unrelated African American males are significantly associated to prostate cancer and related physiological biomarkers.

BODY

The specific aims as listed in the Statement of Work are as follows:

Task1. Start-up phase and subject recruitment (Months 1-5).

- Recruit and hire a research associate.
- Identify and recruit subjects into study.
- Evaluate clinical status of subjects.

Task 2. Data collection (Months 3-20).

- Extraction of genomic DNA from blood samples.
- Genotyping of DNA samples.
- Collection of Epidemiological data.

Task 3. Interim analyses (Months 10-22).

- Infer haplotypes from genotypic data.
- Enter genetic data and epidemiological data into database.
- Perform preliminary data analysis.

Task 4. Final analyses, publications and presentations (Months 18-24).

- Perform data and statistical analyses.
- Test hypotheses.
- Manuscript preparations.

TASK 1:
Dr. Kittles has established individual collaborations with Srinivasan Vijayakumar, M.D., a radiation oncologist, from the University of Chicago and Michael Reese Hospital in Chicago; Sally Weinrich, Ph.D. at University of South Carolina, Columbia, SC; and Chiledum A. Ahaghotu, MD., a urologist at Howard University Hospital. Over the past year, the collaborators have been quite successful in recruiting a cohort of cases and controls from the African American and European American communities in Columbia, SC, Chicago and Washington, DC. The table below details the numbers recruited thus far.

Table 1.

Population	Prostate cancer patients	Age matched controls	Total
African Americans	510	705	1,215
European Americans	200	300	500
TOTAL	**700**	**1100**	**1,715**

TASKS 2, 3 and 4:
Genomic DNA has been extracted from all blood specimens collected using a slight variation of the Puregene DNA extraction protocol. We have started genotyping the androgen receptor gene trinucleotide repeat polymorphisms. The androgen receptor (AR) interacts with androgens to promote cell division (normal and malignant) in the prostate gland. The AR binds dihydrotestosterone and stimulates the transcription of a cascade of androgen responsive genes. Because of this relationship, it has been proposed by many that the AR may be one genetic predictor of susceptibility to prostate cancer. There are two polymorphic regions in the N-terminal protein domain of the AR, which are encoded in the first exon of the AR gene. These are the polyglutamine repeat region (CAG)n and the polyglycine repeat region (GGC)n (Stanford et al., 1997; Irvine et al., 1995; Giovannucci et al., 1997; Edwards et al., 1992). We genotyped the CAG and GGC loci for approximately 950 individuals (1063 chromosomes) using florescent-dye labeled PCR primers and the ABI 377 DNA sequencer. Control populations we have examined thus far for the CAG and GGC markers include African Americans (N=520), Gold Coast Africans (Nigeria and Ghana, N=85), Sierra Leoneans (N=210), European Americans (N=85), Amerindians (N=103), and Asians (Chinese, N=60).

A total of 27 CAG (range 5-31 repeats) and 23 GGC alleles (range: 2-24 repeats) were observed. Not surprisingly, African Americans had the most alleles. The European and Asian populations possessed the least number of alleles at the two loci. Populations of African descent possessed significantly shorter repeats than non-African populations (paired t-test, $p<0.00001$). The entire range of CAG repeat variation was observed among

populations of African descent. Interestingly, the purported high-risk CAG repeat lengths <20, were most prevalent among populations of African descent. The non-random association of CAG and GGC alleles, linkage disequilibrium (LD), was assessed for each study population. Significant evidence for linkage disequilibrium between the two markers was observed only among African Americans cases and controls (p=0.00001) and Amerindians (p=0.009). The LD observed in Amerindians was consistent with their population history of recent population bottlenecks. The high level of linkage disequilibrium among African Americans is likely due to admixture. This assessment of linkage disequilibrium in the African American population is quite significant for several reasons. First, the high level of stratification in the African American population may be a confounder in disease association studies if the substructure is not controlled for. Secondly, the identification of high-risk haplotypes is potentially more powerful in disease studies than single locus analyses. A preliminary analysis of androgen receptor haplotype risk and prostate cancer has revealed an association of closely related haplotypes with high-grade cancer. We intend to increase the resolution in identification of these possible high-risk haplotypes by typing single nucleotide polymorphisms (SNPs) within the gene. This work was published in *Human Genetics* in late 2001.

Another gene we have studied is the human steroid 5α-reductase type 2 gene (SRD5A2) located on chromosome 2. SRD5A2 encodes the isoenzyme 5α-reductase, which is responsible for the intracellular conversion of testosterone to its reduced form, dihydrotestosterone (DHT). DHT promotes prostate cell division and may be involved in benign and neoplastic growth of the prostate in elderly men (Labrie et al., 1993). It has also been suggested that differences in androgen synthesis and metabolism may be responsible for ethnic variation in prostate cancer risk (Ross et al., 1992). Thus genetic variability of the SRD5A2 gene and subsequent enzyme activity may be important risk factors in prostate cancer. A dinucleotide repeat (TA) marker has been observed in exon 5 of the gene. Preliminary studies have shown that, like the androgen receptor CAG and GGC repeat loci, allelic distributions of this polymorphic marker vary considerably between high-risk and low-risk populations (Reichardt et al., 1995). Similarly to the androgen receptor, the TA-repeat and a SNP which creates the loss of an RsaI restriction site within the SRD5A2 gene has been typed for all the samples collected thus far. In addition we have started screening the entire gene for SNPs using a core set of DNA samples from our cohort. Exon 1 has been screened and sequenced for about 60 samples of men with prostate cancer, 10 African control samples, and 25 Asians. We are currently characterizing a SNP in this exon, which contributes to a loss of a BstUI site. We genotyped this SNP using florescent labeled primers and the ABI 377 sequencer along with two other markers, the (TA) repeat located in exon five, and an RsaI RFLP in order to create haplotypes. Haplotype analyses (cladistic analyses) revealed no association with prostate cancer or related clinical phenotypes.

We also examined another genetic region that has been shown to play a role in hereditary forms of prostate cancer. This region is on chromosome X. Evidence for a prostate cancer susceptibility locus on the X chromosome has been observed using linkage analysis on certain families by NHGRI investigators (Xu et al., 1998). The region implicated, Xq27-28 is not near the androgen receptor. In fact more than 50cM separates the suspected

locus from the androgen receptor. Five microsatellites on chromosome X near the Xq27-28 region were genotyped using florescent-labeled primers and the ABI 377 sequencer. The microsatellites included DXS8106, DXS984, DXS1193, DXS1205, and DXS1227. Allele and haplotype analyses of this region on chromosome X did not reveal any correlations with prostate cancer in our populations.

Single nucleotide polymorphisms (SNPs) are useful genetic markers to investigate susceptible genes to diseases and drug responsiveness. In order to efficiently genotype SNPs within the candidate genes of interest our laboratory utilized a novel methodology called Pyrosequencing. The method involves immobilization of amplified, biotinylated DNA sequence products which contain a sequence variant onto sepharose beads. Primer extension reactions are carried out by stepwise elongation of the sequencing primer strand, upon sequential addition of different deoxynucleoside triphosphates and simultaneous degradation of unincorporated nucleotides by the enzyme apyrase. As the mini sequencing reaction continues, the complementary DNA strand extends and the DNA sequence is determined from the single peaks in the pyrogram. The Pyrosequencing system is an automated DNA sequence analysis machine (PSQ 96) that uses 96-well microtiter plates, reagents for SNP detection and the accompanying software creates a sequence database of the SNP and also predicts genotype results with quality assessment of individual samples. We recently published a paper on this methodology in the journal *American International Biotechnology Laboratory*.

Cladistic association analyses of androgen receptor haplotypes

In addition to the two microsatellites (CAG and GGN) within exon 1 of the androgen receptor gene there exists a single nucleotide polymorphism (G633A). The SNP is a synonymous substitution that does not appear to alter protein function. However since it is polymorphic it provides information to differentiate haplotypes in order to find possible susceptibility haplotypes for disease. In this study we evaluated if increased risk for prostate cancer and associated clinical characteristics existed among variants of the androgen receptor gene (haplotypes) using the evolutionary history of the haplotypes. Prostate cancer patients were 40-100 years old and histologically diagnosed within the last 2 years. Gleason grade and tumor stage characteristics were combined to define low and high stage/grade. Healthy controls had normal DREs and PSA <4.0 ng/ml. African Americans (107 Pca patients and 165 healthy volunteers) were recruited from Howard University Hospital, Washington, DC. Nigerians (65 prostate cancer patients and 48 healthy controls all belonging to the Yoruba ethnic group) were recruited at University College Hospital in Ibadan and Central Hospital in Benin City, Nigeria. European Americans (121 Pca patients) were recruited at Michael Reese Hospital Chicago, IL. The two tables below provide clinical and genetic information on the populations.

Table 2.

Population (N)	% affected	age range (mean)	grade/stage	PSA (mean)
Nigerians (113)	58%	37-96 (62)	87% high	0.1- 28 (1.68)
African Americans (272)	39%	37-88 (59)	43% high	0.1- 253 (8.17)
European Americans (121)	100%	61-77 (62)	100% low	0.0- 45 (7.32)

Table 3.

Population (N)	CAG (mean)	GGC (mean)	A-allele	# of haplotypes
Nigerians (113)	3-23 (13)	5-18 (13)	66%	59 distinct
African Americans (272)	6-27 (14)	5-22 (13)	57%	102 distinct
European Americans (121)	6-22 (15)	8-19 (14)	16%	58 distinct

Contingency table analyses revealed a strong association with specific androgen receptor haplotypes and prostate cancer.
1. Strong association of AR haplotypes with prostate cancer in African Americans ($p = 0.0006$ and Nigerians ($p = 0.005$).
2. Nominal association with grade/stage in African Americans ($p = 0.05$)

Two-way ANOVA analyses also reveal an association of AR haplotypes with PSA levels in Nigerians ($p = 0.001$). A second round of statistical analyses is being performed in order to confirm our findings.

OTHER CANDIDATE GENES EXAMINED:

It is well known that androgens play an important role in the etiology of prostate cancer. The CYP17 gene encodes the cytochrome P450c17α enzyme which is the rate-limiting enzyme in androgen biosynthesis. A T to C polymorphism in the 5' promoter region has recently been associated with prostate cancer. However contradictory data exists concerning the risk allele. To further investigate the involvement of the CYP17 variant with prostate cancer we typed the polymorphism in three different populations and evaluated its association with prostate cancer and clinical presentation in African Americans. We genotyped the CYP17 polymorphism in Nigerian (n = 56), European American (n = 74), and African American (n = 111) healthy male volunteers along with African American men affected with prostate cancer (n = 71), using Pyrosequencing. Genotype and allele frequencies did not differ significantly across the different control populations. African American men with the CC CYP17 genotype had an increased risk of prostate cancer [odds ratio (OR), 2.8; 95% confidence interval (CI) = 1.0-7.4] compared to those with the TT genotype. A similar trend was observed between the homozygous variant genotype in African American prostate cancer patients and clinical presentation. The CC genotype was significantly associated with higher grade and stage of prostate cancer (OR, 7.1; 95% CI, 1.4-36.1). The risk did not differ significantly by family history or age. Our results suggested that the C allele of the CYP17 polymorphism is significantly associated with increased prostate cancer risk and clinically advanced disease in African Americans. This work was published in *Cancer Epidemiology Biomarkers and Prevention* in 2001.

An A/G SNP within the promoter of the CYP3A4 gene has previously been associated with prostate cancer in African Americans. However, the SNP exhibits large differences in allele frequency between populations. Given that the African American population is genetically heterogeneous because of its African ancestry and subsequent admixture with European Americans, case-control studies using African Americans are highly susceptible to spurious associations. To test for association with prostate cancer we genotyped CYP3A4-V in prostate cancer patients and age and ethnicity matched controls representing African Americans, Nigerians, and European Americans. To detect population stratification among the African American samples 10 unlinked genetic markers were genotyped. To correct for the stratification, as proposed by Reich and Goldstein (2001), the uncorrected association statistic was divided by the average of association statistics across the 10 unlinked markers. Sharp differences in CYP3A4-V frequencies were observed between Nigerian and European American controls (0.87 and 0.10, respectively; $P<0.0001$). African Americans were intermediate at 0.66. An association uncorrected for stratification was observed between CYP3A4-V and prostate cancer in African Americans ($p=0.007$). An association was also observed among European Americans ($p=0.02$) but not Nigerians. In addition, the unlinked genetic marker test provided strong evidence of population stratification among African Americans. Due to the high level of stratification, the corrected P-value was not significant ($p=0.25$). Follow-up studies on a larger dataset will be needed to confirm if the association is indeed spurious, however these results reveal the potential for confounding of association studies using African Americans and the need for study designs that take into account

substructure due to differences in ancestral proportions between cases and controls. This work was recently published (2002) in *Human Genetics*.

KEY RESEARCH ACCOMPLISHMENTS

- Recruitment of 1800 clinically evaluated prostate cancer patients and healthy volunteers.
- Collected over 6,500 genotypes from 3 markers within the androgen receptor gene.
- Collected over 5,200 genotypes from the steroid 5α-reductase type 2 gene.
- Genotyped 5 microsatellites within and around the chromosome Xq27-28 region.
- Developed new genotyping methodology for high-throughput analyses.
- Published seven manuscripts related to the research.

REPORTABLE OUTCOMES

Published manuscripts:

Kittles R, et al. (2001) Extent of linkage disequilibrium between the Androgen Receptor gene CAG and GGC repeats in human populations: Implications for prostate cancer risk. *Human Genetics.* 109:253-261.

Kittles R, Panguluri R, Chen W, Massac A, Ahaghotu C, Jackson A, Ukoli F, Adams-Campbell L, Isaacs W, Dunston G. (2001) CYP17 promoter variant associated with prostate cancer aggressiveness in African Americans. *Cancer Epidemiology Biomarkers and Prevention.* 10:943-947.

Kittles R and Williams R. (2001) Genetics and Cancer: Challenges at the Millennium. *Journal of Registry Management.* 28(3):132-133.

Kittles R, and Adams-Campbell L. (2002) Race, Genetics and Cancer: Consequences of "Racializing" Disease. *ONCOLOGY.* 16(1): 113-118.

Pfaff C, Kittles R, and Shriver M. (2002) Adjusting for Population Structure in Admixed Populations. *Genetic Epidemiology.* 22:196-201.

Kittles R, et al. (2002) Association of CYP3A4 variant with prostate cancer in African Americans is due to population stratification. *Human Genetics.* 110:553-560.

Panguluri R, Chen WD, Kittles, R. (2002) SNP genotyping of candidate genes for complex disease using Pyrosequencing. *American International Biotechnology Laboratory.* 20(7): 30-34.

Manuscripts in preparation:

Kittles R, et al. Coalescence approach to the association of androgen receptor haplotypes and prostate cancer risk.

List of Personnel:

Rick Kittles, Ph.D.
Nadeje Sylvester
Kapil Pangulari, Ph.D.
Robin Satcher
Chiludum Ahaghotu, M.D.
Sally Weinrich, Ph.D

CONCLUSIONS

The African American population possesses unique genetic features due to its history of antiquity and admixture. When a disease such as prostate cancer manifests variation in incidence and mortality between populations, admixed populations provide a population based approach to evaluate the relative importance of genetic factors. The genetic resources generated by this project are directed towards this end and enabled us to utilize genomic technologies to characterize the functional implications of DNA variation in these populations. Since this study takes advantage of the genetics of unrelated men from diverse ethnic populations, the results may be generalizable to the larger American population. The assessment and publication of genetic variation within the candidate genes for prostate cancer is quite significant because it (1) provides accessibility to the data and allow others to compare their data on other populations for the same markers; (2) encourages others to study the same markers in other populations so that their populations can be placed into a global framework; and (3) stimulates researchers to develop new models and methods to analyze the data.

Hum Genet (2001) 109:253–261
DOI 10.1007/s004390100576

ORIGINAL INVESTIGATION

**Rick A. Kittles · Dale Young · Sally Weinrich
Julie Hudson · George Argyropoulos · Flora Ukoli
Lucile Adams-Campbell · Georgia M. Dunston**

Extent of linkage disequilibrium between the androgen receptor gene CAG and GGC repeats in human populations: implications for prostate cancer risk

Received: 26 December 2000 / Accepted: 2 July 2001 / Published online: 4 August 2001

Abstract While studies have implicated alleles at the CAG and GGC trinucleotide repeats of the androgen receptor gene with high-grade, aggressive prostate cancer disease, little is known about the normal range of variation for these two loci, which are separated by about 1.1 kb. More importantly, few data exist on the extent of linkage disequilibrium (LD) between the two loci in different human populations. Here we present data on CAG and GGC allelic variation and LD in six diverse populations. Alleles at the CAG and GGC repeat loci of the androgen receptor were typed in over 1000 chromosomes from Africa, Asia, and North America. Levels of linkage disequilibrium between the two loci were compared between populations. Haplotype variation and diversity were estimated for each population. Our results reveal that populations of African descent possess significantly shorter alleles for the two loci than non-African populations ($P<0.0001$). Allelic diversity for both markers was higher among African Americans than any other population, including indigenous Africans from Sierra Leone and Nigeria. Analysis of molecular variance revealed that approx. 20% of CAG and GGC repeat variance could be attributed to differences between the populations. All non-African populations possessed the same common haplotype while the three populations of African descent possessed three divergent common haplotypes. Significant LD was observed in our sample of healthy African Americans. The LD observed in the African American population may be due to several reasons; recent migration of African Americans from diverse rural communities following urbanization, recurrent gene flow from diverse West African populations, and admixture with European Americans. This study represents the largest genotyping effort to be performed on the two androgen receptor trinucleotide repeat loci in diverse human populations.

R.A. Kittles (✉) · D. Young · G.M. Dunston
National Human Genome Center, Howard University,
2041 Georgia Ave., Washington, DC 20060
e-mail: rkittles@howard.edu,
Tel.: +1-202-8067028, Fax: +1-202-9863972

R.A. Kittles · F. Ukoli · L. Adams-Campbell · G.M. Dunston
Cancer Center, Howard University, Washington, DC 20059, USA

R.A. Kittles · G.M. Dunston
Department of Microbiology, College of Medicine,
Howard University, Washington, DC 20059, USA

S. Weinrich · J. Hudson
Population Studies, South Carolina Cancer Center,
University of South Carolina, Columbia, SC 29203, USA

G. Argyropoulos
Department of Medicine/Endocrinology,
Medical University of South Carolina,
Charleston, SC 29403, USA

Introduction

The human androgen receptor (AR) is a ligand-dependent nuclear transcriptional factor that regulates the expression of genes necessary for the growth and development of both normal and malignant prostate tissue. The AR gene is about 90 kb and is located on chromosome Xq11–12. Exon 1 of the gene encodes the N-terminal domain, which controls transcriptional activation of the receptor. Exon 1 also encodes two polymorphic trinucleotide repeats (CAG and GGC), which code for polyglutamine and polyglycine tracts, respectively in the N-terminal domain. In vitro studies have demonstrated an inverse relationship between CAG repeat length and AR transcriptional activation ability (Chamberlain et al. 1999).

Variations in AR CAG repeat length have been associated with a number of genetic diseases. Spinal ataxia 1 (SCA1), Kennedy's disease, and Huntington's disease are examples of AR loss of function disorders that result from expansion in AR CAG repeat length (LaSpada et al. 1991; Orr et al. 1993). In addition, short CAG and GGC repeat lengths have been widely attributed to increased risk of developing prostate cancer (Giovannucci et al. 1997; Hardy et al. 1996; Platz et al. 1998). More specifically, individuals with CAG repeat lengths less than 20 and GGC repeats less than 16 have been associated with increased

risk of developing prostate cancer (Giovannucci et al. 1997; Platz et al. 1998; Stanford et al. 1997). Striking differences in CAG repeat lengths have been observed between populations. Black men tend to have significantly shorter repeats than their white counterparts (Edwards et al. 1992; Irvine et al. 1995; Sartor et al. 1999). These genetic differences may be potentially important in understanding why populations of African descent are more susceptible to developing prostate cancer. African American men have the highest incident rate of prostate cancer of any ethnic group in the United States (Brawley and Kramer 1996). Increasing evidence suggests that prostate cancer is more prevalent in populations of African descent (Glover et al. 1998; Ogunbiyi and Shittu 1999; Osegbe 1997). However, attempts to explain the disparity in risk between populations are limited. Although diet (i.e., fat intake) may help to explain the high prevalence of prostate cancer among African Americans, such an influence may be limited when considering other populations of African descent (i.e., Caribbeans and West Africans) whose diet differ considerably.

Genetic studies on the AR CAG and GGC loci have focused mainly on European American prostate cancer patients and controls. Little is known about AR haplotypic variation, especially among different human populations. Evaluating variation in AR trinucleotide repeat lengths across human populations may provide a better understanding of the ethnic disparity associated with prostate cancer. Studies have not formally evaluated variation and the extent of linkage disequilibrium between the two trinucleotide repeat loci across human populations and in particular among those that may have contributed to the African American gene pool. This would be an important prerequisite to determining if there are subpopulations of disease chromosomes segregating in high-risk groups such as African Americans.

When the occurrence of pairs of specific alleles at different loci on the same haplotype is not independent, the deviation from the independence is termed linkage disequilibrium (LD). LD is a population genetic phenomenon that has been useful for gene mapping efforts. It is usually found in populations for genetic markers that are tightly (close genetic distance) linked and can be generated by mutation, selection, or admixture of populations with different allele frequencies. Generally disequilibrium is dependent on population size, time (generations), and distance between genetic markers. Normally, the greater the distance between markers, the faster the decay of disequilibrium. However, for highly polymorphic markers such as microsatellites, the high mutation rate contributes significantly to randomizing associations of alleles.

The aim of this study was to formally evaluate variation and the extent of linkage disequilibrium between the AR gene CAG and GGC repeat loci in human populations, particularly those of African descent such as African Americans.

The African American population is genetically and culturally heterogeneous due to their unique history in the United States (Jackson 1993). While a significant portion of the African American gene pool originates from Western and Central Africa, other populations have also contributed to the present genetic makeup of the population. To better understand variation within the African American population we included comparative populations representing West Africans (Nigeria and Sierra Leone), European Americans, Chinese, and Amerindians.

Subjects and methods

Unrelated African American men (n=520) were recruited from Columbia, South Carolina, for prostate-specific antigen (PSA) screening over the past 5 years. Nigerians (n=85) representing the Edo (Bini) ethnic group were recruited in the Udo community near Benin City, Nigeria. European American men (n=90) were recruited from the Washington, DC area. The African American, Nigerian, and European American men were recruited as healthy community-based controls for prostate cancer studies. Inclusion criteria were men between 50 and 80 years of age with PSA levels less than 4.0 ng/ml and normal digital rectal examinations. In addition, unrelated men representing the Mende ethnic group from Sierra Leone (n=240), Han Chinese from Taiwan (n=60), and an Amerindian population from a community in the southwestern United States (n=103) were also included. No clinical data or medical history was collected for the Sierra Leone, Chinese, and Amerindian participants. Informed consent for genetic analysis was obtained for all subjects. Individuals of mixed ancestry were not excluded. Genomic DNA was isolated from whole-blood samples using the Puregene (Gentra Biosystems) DNA isolation kit. The trinucleotide repeat CAG and GGC loci were amplified by PCR using 50 ng genomic DNA. Primers used to amplify the CAG locus were 5′-TCC AGA ATC TGT TCC AGA GCG TG-3′ (forward) and 5′-GCT GTG AAG GTT GCT GTT CCT CAT-3′ (reverse). Primers specific for the GGC locus were 5′-CCA GAG TCG CTC GCG ACT ACT ACA ACT TTC C-3′ (forward) and 5′-GGA CTG GGA TAG GGC ACT CTG CTC ACC-3′ (reverse). Florescent dyes 6-FAM and HEX were used to label the forward primers for GGC and CAG respectively. PCR cycling conditions for the CAG locus were 35 cycles of 95°C for 30 s, 60°C for 30 s, and 72°C for 30 s. Conditions for the GGC locus was 25 cycles of 97°C for 30 s, 55°C for 30 s and 72°C for 1 min.

PCR products for both loci were then pooled and electrophoresed on an ABI 377 DNA sequencer, (ABI, Foster City, Calif., USA). Genescan and Genotyper 5.0 programs (ABI) were used to generate fragment sizes and genotypes. Due to limited genomic DNA some samples could not be typed for both loci. Statistical analyses for comparison of repeat length mean, mode, and variance among populations were performed using Origin 5.0 (Microcal Software, Northampton, Mass., USA). Heterozygosities for the two trinucleotide repeats and for the haplotypes were computed as $n(1-\Sigma p_i^2)/(n-1)$, where p_i represents the frequency of the ith allele or haplotype, and where n is the number of chromosomes drawn from the population. Standard errors were obtained by using equation 8.7 in Nei (1987). Standardized pairwise linkage disequilibrium values (D′; Lewontin 1964) were calculated for all pairs of microsatellite alleles observed within each population. The null hypothesis of linkage equilibrium (D′=0) was tested and P values obtained by Fisher's exact test using the Markov chain (Guo and Thompson 1992) implemented by the computer program Arlequin 1.1 (Schneider et al. 1997).

Differences among populations were assessed by use of the hierarchical analysis of molecular haplotype variance (AMOVA; Excoffier et al. 1992; Michalakis and Excoffier 1996) implemented by the Arlequin 1.1 package. AMOVA performs a hierarchic analysis of three genetic-variance components: ΦST, subpopulations relative to the total population; ΦSC, subpopulations relative to continental groups; and ΦCT, continental groups relative to the total population. For the analysis, three groups containing the six populations were defined: (a) populations of African descent

(Nigerians, Sierra Leoneans, and African Americans); (b) European American population; and (c) Asian descent populations (Chinese and Amerindian). The AMOVA assumed a single stepwise mutation model (DiRienzo et al. 1994; Valdes et al. 1993) for the trinucleotide repeat loci. In addition, pairwise genetic distances between populations were computed from ΦST values; D = ΦST/(1–ΦST) (Slatkin 1995). Significance levels of the genetic variance components were estimated by use of 10,000 random-permutation procedures.

Results

Allelic diversity

Tables 1 and 2 show the allelic diversity observed in the six populations for the CAG and GGC markers. African Americans possessed the greatest number of alleles for both markers, which partially may be due to the larger sample size. However, for the CAG locus 18 alleles were observed among the significantly smaller sample of Nigerians. The African American population possessed the highest gene diversity for the CAG marker of any of the other populations, while greater diversity was observed for the GGC locus in the Sierra Leone population. The number of CAG alleles observed ranged from 11 for Euroamericans to 21 for African Americans (Table 1). For the GGC locus the number of alleles ranged from 4 for both Asians and Amerindians to 17 for African Americans (Table 2). The GGC allele with 15 repeats was highly frequent in non-African populations, ranging from 55% to 80%. The 15-repeat allele was less frequent among West Africans (5–10%) and intermediate in frequency among African Americans at 23%. Strikingly low diversity was observed at the GGC locus for Chinese and Amerindians. Gene diversity for the two populations of Asian ancestry was almost one-third of that observed for the African populations (Table 2).

CAG and GGC allelic distributions are shown in Fig. 1. These distributions portray a shift in the most common allele among African versus non-African populations. Allelic distributions were either unimodal or bimodal for all populations except the Nigerians and Amerindians (Fig. 1). The multimodal CAG allele distribution for the Nigerian and Amerindian populations may be due to genetic drift. For example, among Nigerians the 17 allele at the CAG locus is rare (<0.01), unlike in the other African populations. Among the Amerindians both the 20 and 22 CAG alleles are common while the 19 and 21 alleles are less frequent (Fig. 1).

As an alternative measure of intrapopulation diversity for the microsatellite markers we calculated variances in allele sizes for each locus (Tables 1, 2). Again, genetic drift operating within the Nigerian and Amerindian populations may have contributed to the higher variance in number of CAG repeats than in the other populations. Interestingly, the same trend was not observed in the Nigerian or Amerindian populations for the GGC locus (Table 2). The lowest variance for CAG allele size was observed among the Europeans. Europeans possessed about one-fourth the variance in CAG allele size than among Nigerians.

Variance in allele size for the GGC marker was one-third that of CAG allele size variance. The populations with the lowest GGC allele size variance were Asians and Amerindians. A notable trend observed among the variances calculated was that variances for the African populations were almost twice that of the non-African populations. This is consistent with the findings of other studies that have examined microsatellite diversity. These studies reveal higher gene diversity among African populations and significant genetic differences between African and non-African populations (Jorde et al. 1995, 2000; Nei and Takezaki 1996; Reich and Goldstein 1998; Shriver et al. 1997).

Table 1 CAG allelic diversity (*N* number of chromosomes, *H* gene diversity)

Population	N	H	No. of alleles	Mean	Range	Variance
African American	516	0.951	21	17.8	9–31	10.97
Sierra Leone	230	0.918	17	17.3	10–26	7.77
Nigerian	83	0.909	18	16.7	5–28	17.28
Euroamerican	87	0.866	11	19.7	13–26	5.37
Asian	60	0.846	12	20.1	14–26	4.55
Amerindian	80	0.884	14	20.1	14–30	8.62

Table 2 GGC allelic diversity (*N* number of chromosomes, *H* gene diversity)

Population	N	H	No. of alleles	Mean	Range	Variance
African American	472	0.880	17	14.3	4–20	4.94
Sierra Leone	210	0.906	14	13.7	4–24	5.50
Nigerian	78	0.771	10	13.8	8–19	3.44
Euroamerican	80	0.628	13	15.0	2–20	5.77
Asian	60	0.322	4	14.6	10–16	1.48
Amerindian	103	0.362	4	14.6	8–16	1.91

Haplotype diversity and LD

AR CAG and GGC haplotype frequencies and D′ values were determined for each population and are available from the Human Genome Diversity Laboratory. Haplotype diversity was greatest for African populations, lower for European Americans, and lowest for Amerindians (Table 3). The 15 most common AR haplotypes, their frequencies, and LD values are shown in Table 4. Highly divergent haplotypes were observed among the African populations. All non-African populations possessed the same most common haplotype designated as 20-15 (20 CAG repeats and 15 GGC repeats). Table 4 reveals that the frequency of the 20-15 haplotype in non-African populations ranged from 13% among Euroamericans to 20% among the Asians. Populations of African descent each possessed a different common haplotype. The most common haplotype among African Americans was 16-16 at 5% frequency. Among Nigerians three common haplotypes were

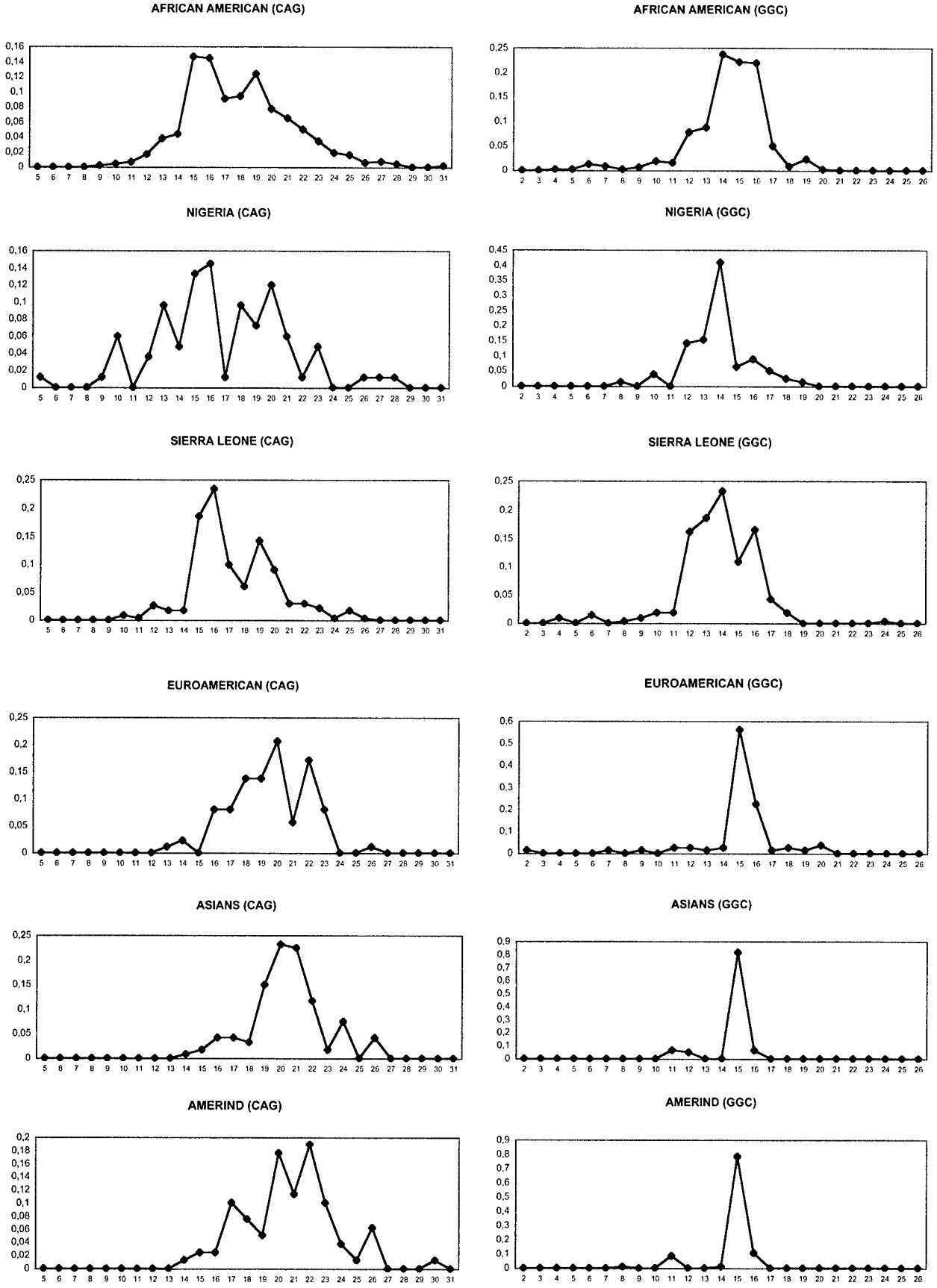

Fig. 1 Allele frequency distributions of CAG (*left*) and GGC (*right*) microsatellites from the six human populations included in this study. *X-axis* Number of repeats; *y-axis* frequency

Table 3 Androgen receptor haplotype diversity (*h*)

Population	No. of haplotypes	h
African American	132	0.991±0.015
Sierra Leone	92	0.990±0.002
Nigerian	50	0.992±0.003
Euroamerican	40	0.980±0.007
Asian	23	0.946±0.018
Amerindian	29	0.938±0.020

observed, 13-14, 19-14, and 20-14, each at a frequency of 6.5%. In Sierra Leone the most common haplotype was 16-12 at 7.4%. The non-African populations did not possess population specific haplotypes. In fact the non-African populations possessed a subset of the variation observed among the African populations. This is similar to observations using other genetic markers (Tishkoff et al. 1996, 1998).

Of all possible pairwise LD comparisons of polymorphic alleles at the two loci, 10% (35 of 357) were significant ($P<0.05$) for African Americans. The percentage of significant D′ values for comparisons of alleles among the other populations ranged from 4% for the Chinese to 8% for Amerindians and European Americans (data not shown). These differences in levels of LD could be due to recent admixture, drift, substructure, or power to detect allelic associations.

Although each non-African population shared the same common haplotype (20-15), no LD was detected between the alleles in the three populations (Table 4). This is likely due to the random effects of drift operating differently in populations after the expansion out of Africa and subsequent mutations away from the common haplotype. Drift and mutation would affect each population differently. This could explain why AR haplotype diversity is greater for the European population than populations of Asian ancestry (Table 3). Our inability to detect LD may also be due to the smaller sample sizes of the non-African populations than the African populations. Sample size likely played a role for the Chinese samples. For instance, D' for many of the common haplotypes in China was 100% but not significant (Table 4).

Table 4 also reveals significant sharing of haplotypes in the two West African populations indicative of shared ancestry. For instance, haplotype 19-14 is common in both Nigeria and Sierra Leone and in strong LD in both populations. This is also reflected in the African American population, where about 40% of the significantly associated alleles were low in frequency (<0.05 allele frequency) and appeared to be of African origin. Table 4 reveals five of these haplotypes (13-14, 15-12, 15-15, 16-12, and 16-16). This is contrary to what would be expected if the LD within the African American population were due to admixture with Euroamericans.

Table 4 Levels of linkage disequilibrium for the 15 most common androgen receptor haplotypes (most common haplotype for each population in *boldface*, *f* haplotype frequency)

Haplotype	African Americans			Nigerians			Sierra Leoneans			European Americans			Asians			Amerindidans		
	f	D' (%)	*P*	*f*	D' (%)	*P*	*f*	D' (%)	*P*	*f*	D' (%)	*P*	*f*	D' (%)	*P*	*f*	D' (%)	*P*
13.14	0.020	37.0	<0.001	**0.065**	**36.0**	**NS**	–	–	–	–	–	–	–	–	–	–	–	–
15.12	0.002	–80.0	0.01	0.010	13.0	NS	0.042	6.0	NS	–	–	–	–	–	–	–	–	–
15.15	0.040	12.0	0.01	0.010	11.0	NS	0.030	7.0	NS	–	–	–	0.020	100.0	NS	–	–	–
15.16	0.040	8.0	NS	0.052	52.0	<0.001	0.037	2.0	NS	–	–		–	–	–	–	–	–
16.12	0.024	18.0	<0.001	0.010	23.0	NS	**0.074**	**27.0**	**<0.001**	–	–	–	–	–	–	–	–	–
16.16	**0.050**	**20.0**	**<0.001**	0.010	2.0	NS	0.047	5.0	NS	0.030	22.0	NS	–	–	–	–	–	–
17.15	0.010	–30.0	NS	–	–	–	0.005	–53.0	NS	0.050	56.0	NS	0.050	100.0	NS	0.090	48.0	NS
18.15	0.010	–43.0	NS	–	–	–	0.005	–6.0	NS	0.060	–17.0	NS	0.040	100.0	NS	0.030	–34.0	NS
19.14	0.034	3.0	NS	**0.065**	**100.0**	**<0.001**	0.069	30.0	<0.001	0.020	42.0	NS	–	–	–	–	–	–
19.15	0.020	–16.0	NS	–	–	–	0.030	12.0	NS	0.060	–17.0	NS	0.160	100.0	NS	0.040	–1.0	NS
20.14	0.030	15.0	NS	**0.065**	**15.0**	**NS**	0.027	4.0	NS	–	–	–	–	–	–	–	–	–
20.15	0.007	–47.0	NS	–	–	–	0.020	7.0	NS	**0.130**	**9.0**	**NS**	**0.200**	**–2.0**	**NS**	**0.160**	**68.0**	**NS**
21.15	0.020	7.0	NS	0.010	14.0	NS	–	–	–	0.040	45.0	NS	0.180	–33.0	NS	0.110	53.0	NS
22.15	0.010	16.0	NS	–	–	–	0.005	6.0	NS	0.100	–2.0	NS	0.110	100.0	NS	0.150	–4.0	NS
23.16	0.002	–69.0	NS	–	–	–	–	–	–	–	–	–	–	–		0.040	28.0	NS
Others	0.681			0.703			0.609			0.510			0.240			0.380		

Table 5 Genetic differentiation of populations

Type of comparison	Variance (%)	Φ Statistic	*P*
Among groups	18.5	ΦCT = 0.185	0.01
Among populations within groups	1.2	ΦSC = 0.014	<0.001
Within populations	80.3	ΦST = 0.197	<0.001

Table 6 Pairwise genetic distances based on ΦST (below the diagonal) and their significance levels (above the diagonal)

	African Americans	European Americans	Amer-indians	Nigerians	Chinese	Sierra Leoneans
African Americans	–	0.000	0.000	0.019	0.000	0.029
European Americans	0.103	–	0.000	0.000	0.416	0.000
Amerindians	0.224	0.049	–	0.000	0.089	0.000
Nigeria	0.027	0.213	0.338	–	0.000	0.297
China	0.137	–0.002	0.018	0.262	–	0.000
Sierra Leone	0.011	0.212	0.351	–0.001	0.265	–

Analysis of molecular variance

Genetic variance (Φ) statistics for the AR trinucleotide repeat data are shown in Table 5. Using both molecular AR haplotypic differences based on microsatellite repeat-length and haplotype frequencies, AMOVA revealed that the AR trinucleotide repeat diversity is nonrandomly distributed across populations. The amount of genetic variance between the six populations was 19.7% ($P<0.001$). The bulk of genetic variance for the AR gene (80.3%) could be explained by individual differences within populations. The **ΦCT** estimate was 0.185, revealing that 18.5% of the genetic variance was due to differences between the African, Asian, and European descent groups (Table 5).

Pairwise genetic distances between the populations are provided in Table 6. The lowest pairwise distances (≤0.05) between populations were observed among the closely related African descendant populations (African Americans, Nigerians, and Sierra Leone) and between the European and Asian populations (Chinese and Amerindian). High genetic distance values (>0.30) were observed between divergent populations such as Amerindians and West Africans from Nigeria or Sierra Leone (Table 6). A moderate distance value of 0.10 between African Americans and European Americans suggests a shared biohistory. All but three of the population pairwise distances were significant ($P<0.05$). The nonsignificant values reflect the close genetic affinities of the three pairs of populations (Table 6).

Discussion

In order to evaluate the extent of variation and linkage disequilibrium between two trinucleotide repeat loci within the AR gene we typed alleles from both markers in six diverse human populations. Populations of African descent exhibited the highest gene diversity among the populations sampled. The African American population contained more alleles and higher gene diversity than even the indigenous West African populations from Sierra Leone and Nigeria. Asians possessed the lowest gene diversity of all populations and also contained the lowest frequency of "high-risk" short alleles for prostate cancer.

Patterns of allelic variation differed substantially between the six populations. Our data revealed that 80% of men of African decent possessed CAG alleles shorter than 20 repeats while only 50% of non-African men had these short alleles (see Fig. 1). The pattern was more pronounced for the GGC locus, where 50% of African men had GGC alleles shorter than 14 while no more than 13% of men with European and Asian ancestry possessed the short GGC alleles. Thus a greater proportion of the haplotypes defined by short alleles at both loci (<20 CAG and <14 GGC) appear to segregate in African populations than in Asian and European populations. In fact, the African American population possesses a mixture of short allele haplotypes from different African populations. This has never been explored and is quite significant since both the CAG and GGC repeat loci influence the size of the protein, which subsequently affects transactivation of the receptor. These results parallel the prevalence of prostate cancer in human populations. Populations in which shorter CAG and GGC alleles are common, such as Africans, and specifically African Americans, have the highest incidence of prostate cancer in the world. The other end of the ethnic spectrum of prostate cancer incidence reveals that prevalence among Asians, who possess larger trinucleotide alleles, may be up to 50-fold less (Ross et al. 1996). Along with other genetic and environmental factors, this could likely yield a stronger predisposition among the African American population for prostate cancer. Recently a relationship was reported between serum PSA levels and polymorphisms in the PSA and AR genes (Xue et al. 2001). Specifically, serum PSA levels increased by 7% with each decreasing AR CAG repeat allele size among individuals homozygous for a single nucleotide polymorphism in the PSA gene promoter.

Our calculation of variance in the number of trinucleotide repeats provided a reliable measure of diversity since the two markers are microsatellites that conform to a stepwise mutation model (DiRienzo et al. 1994; Valdes et al. 1993). Larger variances and higher numbers of alleles were observed for the CAG locus than the GGC locus

among all six populations. The higher diversity at the CAG locus is likely due to a higher mutation rate at the CAG locus than the GGC locus. CAG repeat variation has been shown to cause several human diseases, such as Kennedy's disease (MIM 313200), Huntington's disease (MIM 143100), and several forms of spinocerebellar ataxias: SCA1 (MIM 164400), SCA2 (MIM 183090), SCA3 (MIM 109150), and SCA7 (MIM 164500). All of these CAG repeat loci are polymorphic in normal individuals. However, there appear to be constraints on allele size in populations since disease results when the CAG repeat lengths reach a certain threshold. A recent study of the ERDA1 locus revealed that large CAG repeats are more common among Asian populations, less common in populations of European ancestry, and least common in African populations (Deka et al. 1999). This pattern is very similar to that which was observed in our study of the AR trinucleotide repeats.

To explore population genetic affinities based on the AR gene CAG and GGC repeats we performed an AMOVA. Almost 20% of AR gene variance is attributed to differences between populations. Pairwise genetic distance values were significant for all population pairs except those with shared ancestry, such as between the Asian and Amerindian populations and Nigerian and Sierra Leone populations. Much of the genetic differences between populations may be due to genetic drift. Since the AR gene is X-linked, it is more vulnerable to the effects of drift than similar markers on other autosomes. This is due to a lower recombination rate and smaller effective population size for X-linked markers (approx. three-fourths of non-X-linked autosomal markers). Although the estimate for AR genetic differentiation (ΦST) between populations is higher than non-X-linked autosomal markers, it is not as high as estimates for the haploid systems of mtDNA and the Y chromosome (see Jorde et al. 2000). This is because the effective population size for mtDNA and the Y chromosome is about one-third lower than for the X chromosome.

Studies that have examined the association of alleles at the AR CAG and GGC repeat loci in relation to the development of prostate cancer have been ambiguous. Irvine et al. (1995) examined AR trinucleotide repeat variation in prostate cancer cases and controls from three ethnic groups, Euroamericans, African Americans, and Asians. LD was observed within the mixed group of cases but not within any one ethnic group. Several factors may have led to this observation. The prostate cancer cases consisted of three diverse populations, and therefore it is highly likely that stratification existed when they were pooled together. Also, since the sample sizes of the three groups were low (<50) it is likely that there was not sufficient statistical power to detect LD between the two markers. Later Stanford and colleagues (1997) examined a relatively large sample of Euroamericans. Their sample size of 301 cases and 277 controls failed to detect any LD between the markers within the two groups. A larger study consisting of 582 cases and 794 controls (Platz et al. 1998) revealed significant LD in both the cases and controls. The Platz et al. (1998) finding, using mainly Euroamerican men, was significant only after they pooled alleles for the two markers into categories of fewer than 23, 23, and more than 23 repeats.

It is a general rule that strong disequilibrium indicates that two marker loci are closely spaced. However, it is not always true that two closely spaced markers show disequilibrium. The frequencies of marker alleles and sample size affect the power to detect LD. Also, recombination and/or mutation hotspots could affect LD by increasing the chance that the associated marker allele will change. Not only is haplotype variation shaped by accumulated mutation within haplotypic lineages, it is also fashioned by recombination events among the lineages. It is unlikely that the decay of LD and pattern of variability observed for the AR CAG and GGC defined haplotypes is due to a recombination hotspot between the markers since they are separated by only 1 kb, and recombination on the X chromosome occurs only in women. However, recombination cannot be ruled out, especially since recombination hotspots are more likely in areas of high GC-rich regions in the genome (Eisenbarth et al. 2000). Population history can also have an effect on the extent of LD. Our observation of no detectable LD among the non-African populations may be explained by recent population growth and the high mutation rate at the CAG repeat locus. This is in contrast to the population bottleneck explanation for higher LD levels outside of Africa (Kidd et al. 1998; Tishkoff et al. 1996, 1998). Our finding of significant LD in the African American population is due mainly to gene flow from other populations. Admixture between populations with divergent allele frequencies can generate LD extended beyond 30 cM (Lautenberger et al. 2000). Finally, genetic drift can greatly affect or reinforce existing associations. The role of genetic drift in increasing or decreasing LD may be more significant among the Amerindians since there were a smaller number of alleles observed among the Amerindians than among the other populations. This is consistent with observations of low genetic diversity among Amerindians due to their history of recent population bottlenecks (Kittles et al. 1999; Nei and Roychoudhury 1993; Urbanek et al. 1996).

As previously stated, the high level of linkage disequilibrium observed among African Americans is likely due to multiple sources of admixture. Of the significant allelic associations between the trinucleotide markers in the African American population from South Carolina, 26% appear to have originated from European Americans, while 39% were shared among West African populations from Nigeria and Sierra Leone. These results suggest that the LD generated in African Americans from Columbia, South Carolina, may be due to recent migration of African Americans from diverse rural communities following urbanization, recurrent gene flow from distinct West African populations, and admixture with European Americans. Columbia is the capital of South Carolina and is located in the center of the state. Many African Americans migrated to this region from the coastal Sea Islands and the Low Country (Berkeley, Charleston, Colleton, and Dorchester

counties) during the early 1900s. In the late 1700s the percentage of persons of African origin was quite high in the coastal areas, including the port of Charleston (ranging from 47% to 93%). It also appears that colonial South Carolinians preferred certain African ethnic groups over others as slaves (Littlefield 1981; Morgan 1998). For instance, for a period of time in South Carolina enslaved Africans from Senegambia were preferred over others (Littlefield 1981). This preference was based mainly on the Senegambian's familiarity with rice production, which was the chief crop cultivated in the Carolinas at the time. However, these preferences changed in time along with the changing slave economy in the colonies. The changing trends, along with the relative isolation of the coastal communities of South Carolina likely led to diverse South Carolina African American populations. Subsequently, divergent haplotypes were brought together as people left the rural communities for more urban areas such as Columbia.

This assessment of linkage disequilibrium in the African American population is quite significant for several reasons. First, the high level of stratification in the African American population may be a confounder in disease association studies if the substructure is not controlled for. Second, the identification of high-risk haplotypes is potentially more powerful in disease studies than single locus analyses. We intend to increase the resolution in identification of these possible high-risk haplotypes for prostate cancer by typing single nucleotide polymorphisms within the gene and performing haplotype analyses. Ultimately these studies will provide a better understanding of the role variation within the AR plays in prostate cancer etiology.

Acknowledgements We thank C. Ahaghotu, S.O.Y. Keita, J. Long and C. Rotimi for helpful discussions. We also thank two anonymous reviewers for useful critiques, and W.T. Garvey, and M. Shriver for DNA samples. This work was supported by grants RR03048-13S1 from the National Institutes of Health and DAMD17-00-1-0025 from the Department of Defense.

References

Brawley OW, Kramer BS (1996) Epidemiology of prostate cancer. In: Volgelsang NJ, Scardino PT, Shipley WU, Coffey DS (eds) Comprehensive textbook of genitourinary oncology. Williams and Wilkins, Baltimore

Chamberlain NL, Driver ED, Miesfeld RL (1994) The length and location of CAG trinucleotide repeats in the androgen receptor N-terminal domain affect transactivation function. Nucleic Acids Res 22:3181–3186

Deka R, Guangyun S, Wiest J, Smelser D, Chunhua S, Zhong Y, Chakraborty R (1999) Patterns of instability of expanded CAG repeats at the ERDA1 locus in general populations. Am J Hum Genet 65:192–198

DiRienzo A, Peterson A, Garza J, Valdes A, Slatkin M, Freimer N (1994) Mutational processes of simple-sequence repeat loci in human populations. Proc Natl Acad Sci USA 91:3166–3170

Edwards A, Hammond HA, Jin L, Caskey CT, Chakraborty R (1992) Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. Genomics 12: 241–253

Eisenbarth I, Vogel G, Krone W, Vogel W, Assum G (2000) An isochore transition in the NF1 gene region coincides with a switch in the extent of linkage disequilibrium. Am J Hum Genet 67:873–880

Excoffier L, Smouse P, Quattro J (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application of human mitochondrial DNA restriction data. Genetics 131:479–491

Giovannucci E, et al (1997) The CAG repeat within the androgen receptor gene and its relationship to prostate cancer. Proc Natl Acad Sci USA 94:3320–3323

Glover F, Coffey D, et al (1998) The epidemiology of prostate cancer in Jamaica. J Urol 159:1984–1987

Guo S, Thompson E (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics 57:212–215

Hardy D, Scher H, Bogenreider T, et al (1996) Androgen receptor CAG repeat lengths in prostate cancer: correlation with age of onset. J Clin Endocrinol Metab 81:4400–4405

Irvine RA, Yu MC, Ross RK, Coetze GA (1995) The CAG and GGC microsatellites are in linkage disequilibrium in men with prostate cancer. Cancer Res 55:1937–1940

Jackson FL (1993) Evolutionary and political economic influences on biological diversity in African Americans. J Black Studies 23:539–560

Jorde L, Bamshad M, Watkins W, Zenger R, Fraley A, Krakowiak P, Carpenter K, Soodyall H, Jenkins T, Rogers A (1995) Origins and affinities of modern humans: a comparison of mitochondrial and nuclear genetic data. Am J Hum Genet 57:523–538

Jorde L, Watkins W, Bamshad M, Dixon M, Rickers C, Seielsatd M, Batzer M (2000) The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. Am J Hum Genet 66:979–988

Kidd KK, Morar B, Castiglione CM, Zhao H, Pakstis AJ, Speed WC, Bonne-Tamir B, Lu RB, Goldman D, Lee C, Nam YS, Grandy DK, Jenkins T, Kidd JR (1998) A global survey of haplotype frequencies and linkage disequilibrium at the DRD2 locus. Hum Genet 103:211–227

Kittles R, Bergen AW, Urbanek M, Virkkunen M, Linnoila M, Goldman D, Long JC (1999) Autosomal, mitochondrial, and Y chromosome variation in Finland: evidence for a male-specific bottleneck. Am J Phys Anthropol 108:381–399

LaSpada AR, Wilson A, Lubahn D, Harding A, Fishbeck K (1991) Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. Nature 352:77–79

Lautenberger JA, Stephens JC, O'Brien S, Smith M (2000) Significant admixture linkage disequilibrium across 30 cM around the FY locus in African Americans. Am J Hum Genet 66:969–978

Lewontin RC (1964) The interaction of selection and linkage. I. General considerations: heterotic models. Genetics 49:49–67

Littlefield DC (1981) Rice and slaves: ethnicity and the slave trade in colonial South Carolina. University of Illinois Press, Champaign

Michalakis Y, Excoffier L (1996) A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. Genetics 142:1061–1064

Morgan P (1998) Slave counterpoint: black culture in the eighteenth century Chesapeake and Lowcountry. University of North Carolina Press, Raleigh

Nei M, Roychoudhury K (1993) Evolutionary relationships of human populations on a global scale. Mol Biol Evol 10:927–943

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Nei M, Takezaki N (1996) The root of the phylogenetic tree of human populations. Mol Biol Evol 13:170–177

Ogunbiyi J, Shittu O (1999) Increased incidence of prostate cancer in Nigerians. J Natl Med Assoc 3:159–164

Orr H, Chubg M, Banifi S, Kwiatkoski T, Servadio A, Beaudet A, McCall A, et al (1993) Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. Nat Genet 4: 221–226

Osegbe D (1997) Prostate cancer in Nigerians: facts and non-facts. J Urol 157:1340

Platz E, Giovannucci E, Dahl D, Krithivas K, Hennekens C, Brown M, Stampfer M, Kantoff P (1998) The androgen receptor gene GGN microsatellite and prostate cancer risk. Cancer Epidemiol Biomarkers Prev 7:379–384

Reich D, Goldstein D (1998) Genetic evidence for a Paleolithic human population expansion in Africa. Proc Natl Acad Sci USA 95:8119–8123

Ross R, Pike M, Coetzee G, Reichardt J, Yu M, Feigelson H, Stanczyk F, Kolonel L, Henderson B (1998) Androgen metabolism and prostate cancer: establishing a model of genetic susceptibility. Cancer Res 58:4497–4504

Sartor O, Zheng Q, Eastham J (1999) Androgen receptor gene CAG repeat length varies in a race-specific fashion in men without prostate cancer. Urology 53:378–380

Schneider S, Kueffer J, Roessli D, Excoffier L (1997) Arlequin ver. 1.1: a software for population genetic data analysis. Genetics and Biometry Laboratory, University of Geneva, Geneva

Shriver M, Jin L, Ferrell R, Deka R (1997) Microsatellite data support an early population expansion in Africa. Genome Res 7: 586–591

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. Genetics 139:457–462

Stanford JL, Just JJ, Gibbs M, Wicklund KG, Neal CL, Blumstein BA, Ostrander EA (1997) Polymorphic repeats in androgen receptor gene: molecular markers of prostate cancer risk. Cancer Res 57:1194–1198

Tishkoff S, et al (1996) Global patterns of linkage disequilibrium at the CD4 locus and modern human origins. Science 271: 1380–1387

Tishkoff S, et al (1998) A global haplotype analysis of the mytonic dystrophy locus: implications for the evolution of modern humans and for the origin of mytonic dystrophy mutations. Am J Hum Genet 62:1389–1402

Urbanek M, Goldman D, Long JC (1996) The apportionment of dinucleotide repeat diversity in Native Americans and Europeans: a new approach to measuring gene identity reveals asymmetric patterns of divergence. Mol Biol Evol 13:943–953

Valdes A, Slatkin M, Freimer N (1993) Allele frequencies at microsatellite loci: The stepwise mutation model revisited. Genetics 133:737–749

Xue WM, Coetzee GA, Ross RK, Irvine R, Kolonel L, Henderson BE, Ingles SA (2001) Genetic determinants of serum prostate-specific antigen levels in healthy men from a multiethnic cohort. Cancer Epidemiol Biomarkers Prev 10:575–579

Hum Genet (2001) 109:262–266
DOI 10.1007/s004390100561

ORIGINAL INVESTIGATION

H. Nagata · H. Mutoh · K. Kumahara · Y. Arimoto
T. Tomemori · D. Sakurai · K. Arase · K. Ohno
T. Yamakoshi · K. Nakano · T. Okawa · T. Numata
A. Konno

Association between nasal allergy and a coding variant of the *FcεRIβ* gene Glu237Gly in a Japanese population

Received: 8 February 2001 / Accepted: 14 June 2001 / Published online: 16 August 2001

Abstract The gene for the β-chain of the high-affinity receptor for IgE (*FcεRIβ*) has been proposed as a candidate gene for atopy. A coding variant Glu237Gly has been studied in various populations with asthma and atopy, and the results were controversial for association of the variant with atopy/asthma. Because nasal allergy is a more common atopic disease and shows less remission than asthma, we analyzed whether the Glu237Gly variant is correlated with nasal allergy. The study enrolled 233 patients with nasal allergy and 100 control subjects. Further, three subgroups were selected: patients with perennial nasal allergy (n=149), Japanese cedar pollinosis (n=189), and allergy to multiple allergens (n=45). The allele frequency of Gly237 in the controls and patients was 0.14 and 0.20, and the frequency of Gly237-positive subjects was 0.23 and 0.356, respectively. There was a significant association between Gly237-positivity and nasal allergy, perennial nasal allergy, Japanese cedar pollinosis, and allergy to multiple allergens. Among all 333 subjects we observed a significant relationship between Gly237 and elevated levels of serum total IgE (>250 IU/ml) and very high IgE (>1000 IU/ml). Among patients positive for a specific IgE, Gly237 was significantly associated with high IgE for house dust, mite, and Japanese cedar pollen. These results suggest that the Glu237Gly variant of the *FcεRIβ* gene is involved in the development of nasal allergy through the process for the production of both specific and nonspecific IgE antibodies.

H. Nagata (✉) · H. Mutoh · K. Kumahara · Y. Arimoto
T. Tomemori · D. Sakurai · K. Arase · K. Ohno · T. Yamakoshi
K. Nakano · T. Okawa · T. Numata · A. Konno
Department of Otorhinolaryngology,
Chiba University School of Medicine,
1-8-1 Inohana, Chuoh-ku, Chiba 260-8670, Japan
e-mail: nagatah@ho.chiba-u.ac.jp,
Tel.: +81-43-2227171, Fax: +81-43-2273442

Introduction

Nasal allergy is a common atopic disease, the prevalence of which is around 20% or more in Western countries (Sly 1999; Strachan et al. 1997). Various factors such as duration of breast feeding, maternal age, social class, heating with wood or coal, and exposure to diesel exhaust fumes are believed to affect the prevalence of nasal allergy (Butland et al. 1997; Duhme et al. 1998). However, it is generally accepted that the best established risk factor for nasal allergy is a family history of allergy, especially nasal allergy (Bahna 1992; Sibbald and Rink 1991; Wright et al. 1994), indicating that genetic factors strongly influence nasal allergy.

Many studies have been performed on the genetics of atopy since Cookson et al. (1989) first described a linkage between serum IgE level and a DNA marker for chromosome 11q in British families. Regarding the chromosome 11q some studies have confirmed the linkage of atopy and bronchial hyperresponsiveness to markers on 11q13 (Adra et al. 1999; Collée et al. 1993; Daniels et al. 1996; van Herwerden et al. 1995; Mao et al. 1997; Shirakawa et al. 1994a; Young et al. 1992), while others have failed to find the linkage (Amelung et al. 1992; Collaborative Study on the Genetics of Asthma 1997; Hizawa et al. 1992; Lympany et al. 1992; Malerba et al. 1999; Ober et al. 1998; Rich et al. 1992; Wjst et al. 1999; Yokouchi et al. 2000). Meanwhile, the gene for the β-chain of the high-affinity receptor for IgE (*FcεRIβ*) has been identified as a candidate gene for this linkage between atopy and 11q13 (Sandford et al. 1993), and two coding variants in exon 6 of *FcεRIβ*, Ile181Leu/Ile183Val and Ile181Leu, are reported to be associated with atopy in British subjects (Shirakawa et al. 1994b). These variants subsequently proved to be rare in other races, and another coding variant Glu237Gly in exon 7 was identified as a more common coding variant (Hill and Cookson 1996). Some reports show an association of this variant with atopy and/or bronchial hyperresponsiveness while others do not. Thus, it appears to await clarification whether this coding variant of *FcεRIβ* influ-

CYP17 Promoter Variant Associated with Prostate Cancer Aggressiveness in African Americans[1]

Rick A. Kittles,[2] Ramesh K. Panguluri, Weidong Chen, Aisha Massac, Chiledum Ahaghotu, Aaron Jackson, Flora Ukoli, Lucile Adams-Campbell, William Isaacs, and Georgia M. Dunston

National Human Genome Center at Howard University [R. A. K., R. K. P., W. C., A. M., G. M. D.], Washington, DC 20059; Division of Urology, Howard University Hospital [R. A. K., C. A., A. J.], Washington, DC 20059; Howard University Cancer Center [R. A. K., F. U., L. A-C., G. M. D.], Washington, DC 20059; and The Johns Hopkins University Cancer Center, Baltimore, Maryland [W. I.]

Abstract

Androgens play an important role in the etiology of prostate cancer. The *CYP17* gene encodes the cytochrome P450c17α enzyme, which is the rate-limiting enzyme in androgen biosynthesis. A *T* to *C* polymorphism in the 5′ promoter region has recently been associated with prostate cancer. However, contradictory data exists concerning the risk allele. To investigate further the involvement of the *CYP17* variant with prostate cancer, we typed the polymorphism in three different populations and evaluated its association with prostate cancer and clinical presentation in African Americans. We genotyped the *CYP17* polymorphism in Nigerian (n = 56), European-American (n = 74), and African-American (n = 111) healthy male volunteers, along with African-American men affected with prostate cancer (n = 71), using pyrosequencing. Genotype and allele frequencies did not differ significantly across the different control populations. African-American men with the CC *CYP17* genotype had an increased risk of prostate cancer (odds ratio, 2.8; 95% confidence interval, 1.0–7.4) compared with those with the TT genotype. A similar trend was observed between the homozygous variant genotype in African-American prostate cancer patients and clinical presentation. The CC genotype was significantly associated with higher grade and stage of prostate cancer (odds ratio, 7.1; 95% confidence interval, 1.4–36.1). The risk did not differ significantly by family history or age. Our results suggest that the *C* allele of the *CYP17* polymorphism is significantly associated with increased prostate cancer risk and clinically advanced disease in African Americans.

Introduction

The incidence of prostate cancer varies significantly across ethnic groups, with African-American men having the highest rates worldwide (1–3). African Americans also appear to present more commonly at an advanced stage with aggressive histology and increased cancer-related mortality (4). Although the more advanced cancers in African Americans may be confounded by social class and access to health care, there is a critical need to explore the etiological pathways (genetic and environmental factors) that contribute to this disparity.

Because the prostate is an androgen-regulated organ, androgens may play a major role in the etiology of prostate cancer. The *CYP17* gene encodes the cytochrome P450c17α enzyme that catalyzes two key steps in the steroid biosynthesis pathway. The first step in the biosynthesis pathway involves the conversion of cholesterol to pregnenolone by *CYP11A1*. Subsequently, pregnenolone is converted to 17α-hydroxypregnenolone and then to dehydroepiandrosterone, a precursor of testosterone, by the P450c17α enzyme. A T to C polymorphism in the 5′ promoter region of the *CYP17* gene has been described (5) which has been associated with increased risk for early-onset familial breast cancer (6–8). Also denoted as the *A2* allele, this single nucleotide polymorphism may create an Sp1-type promoter site. However recent electromobility shift assays have not confirmed Sp1 binding (9).

The *CYP17* gene is a likely candidate for prostate cancer, which, like breast cancer, is hormone-related. To date, four studies have shown an association of the *CYP17* gene and prostate cancer risk, however they have been contradictory in terms of which allele is associated. Two studies, from Sweden and Japan, suggested that the *T* (*A1*) allele was associated with increased risk for prostate cancer (10, 11), whereas independent studies from the United States and Austria reported that the *C* (*A2*) allele confers greater risk (12, 13). It is interesting to note that the *C* (*A2*) allele is more prevalent among Asian populations (8, 11, 12, 14). However, *CYP17* allele and genotype frequencies do not seem to differ between African Americans and European Americans (8, 12, 15), unlike several other candidate genes for prostate cancer which exhibit striking allele frequency differences that parallel differences in prostate cancer incidence (16–18). To date, no allele and genotype frequency data exists on clinically evaluated indigenous Africans and African-American prostate cancer patients. The purpose of this study was to determine whether differences exist in *CYP17* genotype frequencies between African, African-American, and European-American populations and whether the *CYP17* polymorphism was associated with prostate cancer risk in African Americans.

Received 12/20/00; revised 6/29/01; accepted 7/5/01.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] This work was supported by Grants RR03048-13S1 from the NIH and DAMD17-00-1-0025 from the Department of Defense and by funds from the Howard University Cancer Center.

[2] To whom requests for reprints should addressed, at National Human Genome Center at Howard University, 2041 Georgia Avenue NW, Washington, DC 20060. Phone: (202) 806-7028; Fax (202) 986-3972; E-mail: rkittles@howard.edu.

Subjects and Methods

Unrelated men were enrolled from three sites for a population-based study of genetic risk factors for prostate cancer. The Howard University Institutional Review Board approved the study, and written consent was obtained from all subjects. All prostate cancer cases were between 40 and 79 years of age and were diagnosed with prostate cancer within the last 2 years. One hundred and eighty-two African Americans (71 prostate cancer patients and 111 healthy male controls) were enrolled from the Washington, DC area. African-American men with prostate cancer were recruited from the Division of Urology at the Howard University Hospital and/or prostate cancer screening at the Howard University Cancer Center. The response rate among the African-American cases was 92%. Healthy African-American male volunteers were enrolled among individuals undergoing regular physical exams at the Division of Urology at Howard University Hospital and/or men participating in screening programs for prostate cancer at the Howard University Cancer Center. The response rate for the African American controls was 90%. The mean age of prostate cancer patients was 66.3 ± 3.3 years and, among controls, 57.3 ± 0.8 years. Clinical characteristics including Gleason grade, PSA,[3] Tumor-Node-Metastasis stage, age at diagnosis, and family history were obtained from medical records.

Seventy-four European-American healthy male volunteers (mean age, 58.5 ± 2.9 years) were enrolled through various prostate cancer-screening programs in Baltimore, MD, sponsored by the Johns Hopkins Cancer Center. Fifty-six healthy volunteers (mean age, 51.9 ± 1.6 years) belonging to the Edo ethnic group were enrolled in Benin City, Nigeria. Nigerian males were enrolled through a community-based study of risk factors for prostate cancer during the summer of 2000 in collaboration with the University of Benin Teaching Hospital in Benin City, Nigeria. The response rate among the Nigerian controls was 85%. Blood samples were collected from each subject. Ethnicity for all groups was self-reported, and individuals of mixed ancestry were not excluded. All healthy volunteers had PSA levels <4.0 ng/ml and normal digital rectal exams.

Genotyping. The genomic DNA was obtained from isolated lymphocytes using cell lysis, proteinase K-treatment, protein precipitation, and DNA precipitation. Genotyping of the *T* to *C* polymorphism in the promoter region of *CYP17* gene was performed using Pyrosequencing (19, 20). The primers for the polymorphism were designed from the published promoter sequence (National Center for Biotechnology Information accession no. M63871). A 167-bp fragment was amplified in a 50-μl PCR reaction containing 30 ng of genomic DNA, 20 pmol of forward unlabeled 5′-TTC CAC AAG GCA AGA GAT AAC-3′ and a reverse biotin-labeled primer (b indicates biotin) 5′-b-GGT AAG CAG CAA GAG AGC CA-3′ and 1× PCR buffer II (Perkin-Elmer), 2 mM MgCl2, 0.2 mM dNTP, and AmpliTaq gold DNA polymerase. PCR reactions were performed for 50 cycles: denaturation at 95°C for 30 s, annealing at 54°C for 20 s, and extension at 72°C for 30 s.

Biotinylated single-stranded DNA fragments were generated by mixing the PCR product with streptavidin-coated paramagnetic beads (Dynalbeads M280; Dynal, Norway). The PCR products and Dynal beads were mixed with high-salt buffer [0.1% Tween 20, 2 M NaCl, 0.5 mM EDTA, and 10 mM Tris-HCl (pH 7.6)], incubated for 15 min at 65°C, and spun at 1400 rpm in a thermomixer (Eppendorf). Then the material was resuspended in 0.5 M NaOH and incubated for 2 min to separate DNA stands. Dynal beads containing the immobilized strand were washed in 1× annealing buffer (20 mM Tris-Acetate and 5 mM MgAc2) and resuspended in 45 μl of 1× annealing buffer and 10 pmol of sequencing primer. Then the mixture was incubated at 80°C for 2 min and then cooled to room temperature. Throughout the sample preparation steps, the immobilized fragments coupled to Dynal beads were processed using a manifold device (PSQ 96 Sample Preparation Tool; Pyrosequencing AB, Uppsala, Sweden). An automated pyrosequencing instrument, PSQ96 (Pyrosequencing AB) was used to perform genotyping. The reaction was carried out at 25°C with the sequencing primer 5′-GGC AGG CAA GAT AGA CA-3′ added to the reaction. The reaction mixture also contained DNA polymerase (exonuclease-deficient), 40 mU apyrase, 4 μg of purified luciferase/ml, 15 mU of recombinant ATP sulfurylase, 0.1 M Tris-acetate (pH 7.75), 0.5 mM EDTA, 5 mM magnesium acetate, 0.1% BSA, 1 mM DTT, 10 μM adenosine 5′-phosphosulfate, 0.4 mg of poly(vinylpyrolidone)/ml, and 100 μg of D-luciferin/ml. The mini-sequencing protocol was carried out by stepwise elongation of the primer strand upon sequential addition of 40 pmol of the different deoxynucleoside triphosphates and the simultaneous degradation of nucleotides by apyrase. As the sequencing reaction continued, the cDNA strand extended and the DNA sequence was determined from the single peaks in the pyrogram using Pyrosequencing software (Pyrosequencing, AB). All samples were genotyped twice directly from genomic DNA. Control DNAs included a known wild-type (TT), a heterozygous mutant (CT), and homozygous mutant (CC) variant samples. The control DNAs were confirmed by direct DNA sequencing using an ABI 377 DNA sequencer (ABI, Foster City, CA). Genotypes from the repeat assay were 100% concordant with initial genotypes.

Statistical Analysis. Genotype and allele frequencies were calculated for each population. Hardy-Weinberg equilibrium analysis of each group was evaluated by contingency table analysis. The SAS Version 6.12 computer program (SAS Institute, Inc., Cary, NC.) was used to compute the two-sided Pearson χ^2 test. ORs and *P*s were determined from a comparison of genotypes in Nigerians and European Americans *versus* African-American healthy controls. Genotypes were also compared between African-American prostate cancer patients and healthy controls. Regression analyses were used to assess whether age at diagnosis and family history modified the relationship between *CYP17* and prostate cancer risk. Regression analyses were also performed to compare grade/stage among prostate cancer patients. Grade/stage was defined as low (T_{1a}–T_{1c} and/or Gleason grade <7) or high [T_2–T_4 or N (+) or M (+) stage and/or Gleason grade ≥7; see Refs. 16 and 17]. For the analyses of prostate cancer patients, the regression model controlled for age at diagnosis, PSA (total), and family history (affected first-degree relative).

Results

Fig. 1 shows examples of pyrograms representing the *CYP17* genotypes. The *C* (*A2*) allele frequency was 30% among the African-American controls. *CYP17* genotypes frequencies in the three normal control populations are shown in Table 1.

[3] The abbreviations used are PSA, prostate-specific antigen; OR, odds ratio; CI, confidence interval.

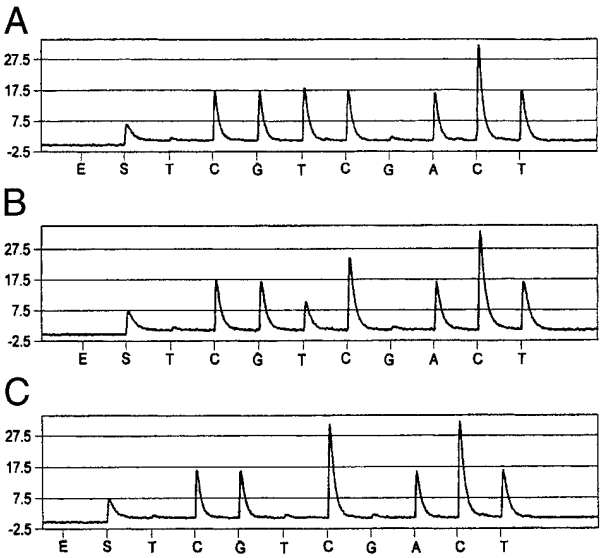

Fig. 1. Pyrograms of the *CYP17* genotypes. The DNA sequence is CGT/CCACCT. *A*, homozygous *TT* sample; *B*, heterozygous *TC* sample; *C*, homozygous *CC* sample. *E* and *S* at the beginning of each pyrogram denote the addition of the enzyme and substrate respectively. The first *T* and second *G* in the programs were negative controls that were used as internal controls for the pyrosequencing reactions.

Table 1 *CYP17* genotype frequencies in healthy controls from three populations

Population	*n*	Genotype *n* (%)			*P*[a]
		TT	*TC*	*CC*	
Nigerians	56	24 (43%)	27 (48%)	5 (9%)	0.69
European Americans	74	28 (38%)	38 (51%)	8 (11%)	0.29
African Americans	111	55 (50%)	46 (41%)	10 (9%)	

[a] Two-sided *P* from Pearson χ^2 tests comparing genotype frequencies in each population with African-Americans.

Genotypes in each population were in Hardy-Weinberg equilibrium ($P > 0.05$; data not shown). Genotype frequencies for European Americans and African Americans were consistent with previous published frequencies (8, 12, 15). Pearson χ^2 tests revealed no significant differences in genotype frequencies when African Americans were compared with Nigerians or European Americans (Table 1).

The presence of at least one copy of the *C* (*A2*) allele was significantly higher among African-American prostate cancer cases (69%), than among controls (50%); $P = 0.01$ (Table 2). An increased risk for prostate cancer was observed for individuals with at least one copy of the *C* allele (OR, 2.2; 95% CI, 1.2–4.1). The risk for prostate cancer among heterozygous individuals was intermediate to those who were homozygous for the *C* allele (ORs of 2.0 and 2.8, respectively; Table 2). This suggests a gene-dosage effect where the risk for prostate cancer increases with number of *C* allele copies. Additional analyses were performed to examine whether a relationship exists between *CYP17* genotype and age at diagnosis (<66 years of age *versus* >66 years of age) and family history of prostate cancer. No relationship was observed between the *CYP17* polymorphism and age of onset in African Americans ($P = 0.71$). Similarly, no association was observed with family history ($P = 0.65$; data not shown).

Table 2 *CYP17* genotype frequencies in African-American prostate cancer cases and healthy controls

Genotype	Cases	Controls	OR	95% CI	*P*[a]
	n = 71	*n* = 111			
TT	22 (31%)	55 (50%)	1.0 (Ref.)		
TC	38 (54%)	46 (41%)	2.0	1.0–3.9	0.03
CC	11 (15%)	10 (9%)	2.8	1.0–7.4	0.04
TC + *CC*	49 (69%)	56 (50%)	2.2	1.2–4.1	0.01

[a] Two-sided *P* from Pearson χ^2 tests.

Table 3 Comparison of *CYP17* genotype with grade/stage[a] among African-American prostate cancer patients

Genotype	Low	High	OR	95% CI	*P*[a]
	n = 37	*n* = 34			
TT	16 (43%)	6 (18%)	1.0 (Ref.)		
TC	18 (49%)	20 (58%)	2.9	1.0–9.2	0.05
CC	3 (8%)	8 (24%)	7.1	1.4–36.1	0.01
TC + *CC*	21 (57%)	28 (82%)	3.6	1.2–10.6	0.01

[a] Grade/stage as defined as low (T_{1a}–T_{1c} stage and/or Gleason grade <7) or high (T_2–T_4 or N (+) or M (+) stage and Gleason grade ≥7).
[b] *P* from logistic regression analyses controlling for age, PSA, and family history.

Stratification of the 71 African-American prostate cancer cases by grade/stage is shown in Table 3. Among men heterozygous for the *CYP17* polymorphism, 58% (20 of 34) presented with high grade/stage prostate cancer compared with 49% (18 of 37) with low grade/stage. For the *CC* genotype, we observed 24% (8 of 37) of men with high grade/stage compared with only 8% (3 of 37) of men with low grade/stage. ORs comparing *TC* genotype to *TT* between low and high grade/stage disease suggests an increased risk of presenting with high grade/stage (OR, 2.9; 95% CI, 1.0–9.2). A stronger association was observed when comparing *CC* genotype to *TT* (OR of 7.1; 95% CI, 1.4- 36.1; $P = 0.01$). Because of the small number of samples in certain categories, the 95% CIs for the ORs are large.

Discussion

Prostate cancer development is influenced by androgens, which are regulated by genetic and environmental factors. Environmental factors such as dietary fat intake play a role in the development of prostate cancer (21). *CYP17* is an ideal candidate for prostate cancer because it is directly involved in the production of testosterone. In this study, we examined the role a *CYP17* promoter polymorphism plays in prostate cancer among African Americans. The *CYP17* polymorphism is in the promoter region and may create an additional Sp1-type site (CCACC) 34 bases upstream of the initiation of translation and downstream from the transcription start site. The presence of this variant may result in increased production of testosterone attributable to an increased rate of transcription (5). This would increase the bioavailability of testosterone for conversion to dihydrotestosterone, ultimately affecting prostate cell growth. Kristensen *et al.* (9) demonstrated that the *CYP17* promoter polymorphism does not create an Sp1 binding site, but suggested that other transcription factors might interact with this polymorphism. However, it is possible that *in vivo* conditions may favor Sp1 binding to the variant Sp1 site in the prostate, thus bringing about increased transcription of the *CYP17* gene.

Two important risk factors for prostate cancer are age and ethnicity. The *CYP17* polymorphism was significantly associated with disease and aggressiveness, and its effect did not seem to be modified by age at diagnosis or family history. The *CYP17* association with prostate cancer among African Americans may also explain the higher circulating testosterone concentration in African-American men when compared with other ethnic groups (18, 22), because the gene is directly involved in testosterone biosynthesis. It is critical to determine whether differences exist in allele and genotype frequency between populations, because this may help explain some of the differences between populations in prostate cancer prevalence. In this study we showed that the frequency of the *CYP17* variant was consistent across control populations consisting of Nigerians, African Americans, and European Americans. This is an important observation, because it suggests that *CYP17* may not account for all of the differences in testosterone levels and prostate cancer incidence between populations. Also, allele frequency differences between populations can be a confounder in association studies if not controlled for (23–25), especially in genetic studies on the African-American population, which is highly heterogeneous because of its African ancestry and recent admixture with European Americans.

This is the first study that investigated the relationship between the CYP17 polymorphism and prostate cancer in African Americans. Although the observed association of the *CYP17 C* (*A2*) allele with prostate cancer is consistent with previous studies on Austrians (13) and European Americans from South Carolina (12), no association has previously been shown with clinical presentation of the disease. Also, conflicting results have been published as to the associated allele. The *T* (*A1*) allele was associated with increased prostate cancer risk in the Japanese (11) and the Swedish populations (10). It has been suggested that the *CYP17* genotypes may play either a protective or a promoting role in prostate cancer progression, given different environmental and/or genetic backgrounds (11). Different populations exhibit different environmental factors (diet, lifestyle, etc.), levels of genetic variation, and patterns of genotype/environment interactions. All of these factors play a role in prostate cancer progression. This may be one of several reasons for the contradictory results. Another reason could be that the *T* to *C* promoter polymorphism within the *CYP17* gene is in moderate (or incomplete) linkage disequilibrium with the actual disease-related polymorphism. It is likely that the disease allele is older in age than the promoter polymorphism because both promoter alleles (*T* and *C*) have been found to be associated with the disease in vastly different populations. Events such as recombination could place the disease allele on different *CYP17* haplotypic backgrounds, and so single marker studies would produce conflicting results. This could be evaluated by screening the *CYP17* gene for more polymorphisms, estimating the level of linkage disequilibrium, and performing haplotypic (multisite) association analyses on prostate cancer in different populations.

In summary, a common *CYP17* variant was associated with increased risk of prostate cancer in African-American men. Comparison of genotypes revealed a significantly higher risk among individuals homozygous for the *C* allele for developing high grade/stage prostate cancer. In fact, African-American patients with the *CC* genotype were seven times more likely to present with more aggressive disease. Because the sample sizes were moderate for the African-American samples, the results should be interpreted with caution until larger studies further evaluate the polymorphism. Future research on the role polymorphisms within the *CYP17* gene play in prostate cancer and clinical presentation may demonstrate a need for genetic screening, possibly providing better treatment opportunities or prevention strategies. However, other genes have been identified that also are involved in prostate cancer, and *CYP17* may play a small but important role in the etiology of prostate cancer.

Acknowledgments

We thank all participants for volunteering in this study. We also thank Norma Foster, Nadeje Sylvester, and Dale Young for their assistance in subject enrollment and data and sample processing.

References

1. Brawley, O. W., and Kramer, B. S. *In:* N. J. Volgelsang, P. T. Scardino, W. U. Shipley, and D. S. Coffey (eds.), Comprehensive Textbook of Genitourinary Oncology. Baltimore, MD: Williams & Wilkins, 1996.

2. Miller, B., Kolonel, L., Bernstein, L., Young, J., Swanson, G., West, D., Key, C., Liff, J., Glover, C., and Alexander, G. Racial/Ethnic Patterns of Cancer in the United States 1988–1992. NIH Publ. No. 96-4104. Bethesda, MD, National Cancer Institute, 1996.

3. Parkin, D., Pisani, P., and Ferlay, J. Estimates of the worldwide incidence of 25 major cancers in 1990. Int. J. Cancer, *80:* 827–841, 1999.

4. Robbins, A. S., Whittemore, A. S., and Van Den Eeden, S. K. Race, prostate cancer survival, and membership in a large health organization. J. Natl. Cancer Inst., *90:* 986–990, 1998.

5. Carey, A. H., Waterworth, D., Patel, K., White, D., Little, J., Novelli, P., Franks, S., and Williamson, R. Polycystic ovaries and premature male pattern baldness are associated with one allele of the steroid metabolism gene *CYP17*. Hum. Mol. Gen., *3:* 1873–1876, 1994.

6. Spurdle, A., Hopper, J., Dite, G., Chen, X., Cui, J., McCredie, M., Giles, G., Southey, M., Venter, D., Easton, D., and Chenevix-Trench, G. CYP17 promoter polymorphism and breast cancer in Australian women under age forty years. J. Natl. Cancer Inst., *92:* 1674–1681, 2000.

7. Bergman-Jungestrom, M., Gentile, M., Lundin, A. C., and Wingren, S. Association between *CYP17* gene polymorphism and risk of breast cancer in young women. Int. J. Cancer, *84:* 350–353, 1999.

8. Feigelson, H. S., Coetzee, G. A., Kolonel, L., Ross, R. K., and Henderson, B. E. A polymorphism in the *CYP17* gene increases the risk of breast cancer. Cancer Res., *57:* 1063–1065, 1997.

9. Kristensen, V. N., Haraldsen, E. K., Anderson K. B., Lonning, P. E., Erikstein, B., Karesen, R., Gabrielsen, O. S., and Borresen-Dale, A. L. *CYP17* and breast cancer risk: the polymorphism in the 5′ flanking area of the gene does not influence binding to Sp-1. Cancer Res., *59:* 2825–2828, 1999.

10. Wadelius, M., Anderson, S., Johansson, J., Wadelius, C., and Rane, A. Prostate cancer associated with *CYP17* genotype. Pharmacogenetics, *9:* 635–639, 1999.

11. Habuchi, T., Liqing, Z., Suzuki, T., Sasaki, R., Tsuchiya, N., Tachiki, H., Shimoda, N., Satoh, S., Sato, K., Kakehi, Y., Kamoto, T., Ogawa, O., and Kato, T. Increased risk of prostate cancer and benign prostatic hyperplasia associated with a *CYP17* gene polymorphism with a gene dosage effect. Cancer Res., *60:* 5710–5713, 2000.

12. Lunn, R. M., Bell, D. A., Mohler, J. L., and Taylor, J. A. Prostate cancer risk and polymorphism in 17 hydroxylase (CYP17) and steroid reductase (SRD5A2). Carcinogenesis (Lond.), *20:* 1727–1731, 1999.

13. Gsur, A., Bernhofer, G., Hinteregger, S., Haidinger, G., Schatzl, G., Madersbacher, S., Marberger, M., Vutuc, C., and Micksche, M. A polymorphism in the *CYP17* gene is associated with prostate cancer risk. Int. J. Cancer, *87:* 434–437, 2000.

14. Huang, C. S., Chern, H. D., Chang, K. J., Cheng, C. W., Hsu, S. M., and Shen, C. Y. Breast cancer risk associated with genotype polymorphism of the estrogen-metabolizing genes *CYP17*, *CYP1A1*, and *COMT*: a multigenic study on cancer susceptibility. Cancer Res., *59:* 4870–4875, 1999.

15. Weston, A., Pan, C. F., Bleiweiss, I. J., Ksieski, H. B., Roy, N., Maloney, N., and Wolf, M. S. *CYP17* genotype and breast cancer risk. Cancer Epidemiol. Biomark. Prev., *7:* 941–944, 1998.

16. Paris, P. L., Kupelian, P. A., Hall, J. M., Williams, T. L., Levin, H., Klein, E. A., Casey, G., and Witte, J. S. Association between a *CYP3A4* genetic variant and clinical presentation in African-American prostate cancer patients. Cancer Epidemiol. Biomark. Prev., *8:* 901–905, 1999.

17. Rebbeck, T. R., Jaffe, J. M., Walker, A. H., Wein, A. J., and Malkowicz, S. B. Modification of clinical presentation of prostate tumors by a novel genetic variant in CYP3A4. J. Natl. Cancer Inst., *90:* 1225–1229, 1998.

18. Ross, R. K., Pike, M. C., Coetzee, G. A., Reichardt, J. K. V., Yu, M. C., Feigelson, H., Stanczyk, F. Z, Kolonel, L. N., and Henderson, B. E. Androgen metabolism and prostate cancer: establishing a model of genetic susceptibility. Cancer Res., *58:* 4497–4504, 1998.

19. Ahmadian, A., Gharizadeh, B., Gustafsson, A. C., Sterky, F., Nyren, P., Uhlen, M., and Lundeberg, J. Single-nucleotide polymorphism analysis by pyrosequencing. Anal. Biochem., *280:* 103–110, 2000.

20. Alderborn, A., Kristofferson, A., and Hammerling, U. Determination of single-nucleotide polymorphisms by real-time pyrophosphate DNA sequencing. Genome Res., *10:* 1249–1258, 2000.

21. Giovannucci, E., Rimm, E. B., Colditz, G. A., Stampfer, M. J., Ascherio, A., Chute, C. C., and Willett, W. C. A prospective study of dietary fat and risk of prostate cancer. J. Natl. Cancer Inst., *85:* 1571–1579, 1993.

22. Ross, R. K., Coetzee, G. A., Reichardt, J. K., Skinner, E., and Henderson, B. E. Does the racial-ethnic variation in prostate cancer risk have a hormonal basis? Cancer (Phila.), *75:* 1778–1782, 1995.

23. Lander, E. S., and Schork, N. J. Genetic dissection of complex traits. Science (Wash. DC), *265:* 2037–2048, 1994.

24. Pritchard, J. K., Stephens, M., Rosenberg, N., and Donnelly, P. Association mapping in structured populations. Am. J. Hum. Genet., *67:* 170–181, 2000.

25. Wacholder, S., Rothman, N., and Caporaso, N. Population stratification in epidemiologic studies of common genetic variants and cancer: quantification of bias. J. Natl. Cancer Inst., *92:* 1151–1158, 2000.

Hum Genet (2002) 110:553–560
DOI 10.1007/s00439-002-0731-5

ORIGINAL INVESTIGATION

Rick A. Kittles · Weidong Chen · Ramesh K. Panguluri
Chiledum Ahaghotu · Aaron Jackson
Clement A. Adebamowo · Robin Griffin · Tyisha Williams
Flora Ukoli · Lucile Adams-Campbell · John Kwagyan
William Isaacs · Vincent Freeman · Georgia M. Dunston

CYP3A4-V and prostate cancer in African Americans: causal or confounding association because of population stratification?

Received: 27 February 2002 / Accepted: 29 March 2002 / Published online: 7 May 2002

Abstract *CYP3A4-V*, an A to G promoter variant associated with prostate cancer in African Americans, exhibits large differences in allele frequency between populations. Given that the African American population is genetically heterogeneous because of its African ancestry and subsequent admixture with European Americans, case-control studies with African Americans are highly susceptible to spurious associations. To test for association with prostate cancer, we genotyped *CYP3A4-V* in 1376 (2 N) chromosomes from prostate cancer patients and age- and ethnicity-matched controls representing African Americans, Nigerians, and European Americans. To detect population stratification among the African American samples, 10 unlinked genetic markers were genotyped. To correct for the stratification, the uncorrected association statistic was divided by the average of association statistics across the 10 unlinked markers. Sharp differences in *CYP3A4-V* frequencies were observed between Nigerian and European American controls (0.87 and 0.10, respectively; $P<0.0001$). African Americans were intermediate (0.66). An association uncorrected for stratification was observed between *CYP3A4-V* and prostate cancer in African Americans ($P=0.007$). A nominal association was also observed among European Americans ($P=0.02$) but not Nigerians. In addition, the unlinked genetic marker test provided strong evidence of population stratification among African Americans. Because of the high level of stratification, the corrected *P*-value was not significant ($P=0.25$). Follow-up studies on a larger dataset will be needed to confirm whether the association is indeed spurious; however, these results reveal the potential for confounding of association studies by using African Americans and the need for study designs that take into account substructure caused by differences in ancestral proportions between cases and controls.

R.A. Kittles (✉) · W. Chen · R.K. Panguluri · R. Griffin
T. Williams · J. Kwagyan · G.M. Dunston
National Human Genome Center, Howard University,
2041 Georgia Avenue, Washington, DC 20060, USA
e-mail: rkittles@howard.edu

R.A. Kittles · C. Ahaghotu · A. Jackson
Divison of Urology, Howard University Hospital,
Washington, DC 20060, USA

R.A. Kittles · C. Ahaghotu · F. Ukoli · L. Adams-Campbell
G.M. Dunston
Howard University Cancer Center, Washington, DC 20060, USA

C.A. Adebamowo
University College Hospital, Ibadan, Nigeria

W. Isaacs
Johns Hopkins Oncology Center, Johns Hopkins University,
Baltimore, MD 21231, USA

V. Freeman
Department of Preventive Medicine and Epidemiology,
Strich School of Medicine, Loyola University Medical Center,
Maywood, IL 60153, USA

Introduction

Given the role that androgens play in prostate development, genes involved in androgen biosynthesis and metabolism may be important factors involved in the etiology of prostate cancer. One such gene may be the *CYP3A4* gene, a member of the cytochrome P450 supergene family involved in the oxidative deactivation of testosterone (Waxman et al. 1998). Recently, *CYP3A4-V*, an A to G polymorphism in the nifidipine-specific element (NSFE) of the 5' regulatory region of the gene has been associated with higher Gleason grade and TNM stage (pathologic system of tumor classification) prostate cancer (Rebbeck et al. 1998; Paris et al. 1999). The associations were most pronounced among men older than 65 years of age with no family history (Rebbeck et al. 1998; Paris et al. 1999).

It has been suggested that *CYP3A4-V* decreases CYP3A4 protein activity thus increasing the availability of testosterone (Paris et al. 1999; Rebbeck 2000). Although there is no consensus on a direct functional correlation of the *CYP3A4* polymorphism (Westlind et al. 1999; Amirimani et al. 1999; Ando et al. 1999; Ball et al. 1999), there does

appear to be a strong link between high levels of testosterone and prostate cancer development (Ross et al. 1995, 1998). Interestingly, allele frequencies for the *CYP3A4-V* vary significantly between populations within the US (Rebbeck et al. 1998; Paris et al. 1999; Walker et al. 1998) with a trend similar to prostate cancer incidence. The incidence of prostate cancer is roughly 60% higher among African American men than European American men (Ries et al. 2000). Similarly, *CYP3A4-V* allele frequency is much higher among African Americans than among European Americans and Asians (Rebbeck et al. 1998; Paris et al. 1999; Walker et al. 1998). These observations are important because the African American population is highly heterogeneous because of African ancestry and subsequent admixture with European Americans. This unique population history could have major consequences for association studies when the risk for disease varies considerably between African American and European American populations. If the frequency of a genetic polymorphism also varies between the ethnic groups, then the polymorphism will appear to be related to the disease (Chakraborty and Weiss 1988; Lander and Schork 1994; Wacholder et al. 2000). This confounding resulting from population stratification is of special concern to genetic epidemiologists, especially with the increased attention on single nucleotide polymorphisms (SNPs) to facilitate population-based methods for genetics studies of complex disease (Collins et al. 1997; Kruglyak 1999). Unfortunately, the underlying population structure may not be known, and crude proxies such as "race" may not sufficiently resolve the level of stratification that may exist within the population. Recently, it has been proposed that unlinked genetic markers be typed in order to detect, quantify, and correct for stratification in case-control studies (Pritchard and Rosenberg 1999; Pritchard et al. 2000; Devlin and Roeder 1999; Schork et al. 2001; Reich and Goldstein 2001). The rationale behind the unlinked marker analysis is straightforward. If stratification exists, not only would the candidate marker be associated but also unlinked markers.

Genetic association studies on prostate cancer in African Americans are particularly vulnerable to confounding attributable to population stratification because one of the major risk factors for prostate cancer is ethnicity (Ross et al. 1998; Greenlee et al. 2000). The incidence of prostate cancer varies significantly across ethnic groups, with African American men having the highest rates worldwide; they are almost two times more likely to develop prostate cancer than European American men (Greenlee et al. 2000). The ethnic variation in prostate cancer incidence suggests that genetic factors in combination with environmental factors play a vital role in determining prostate cancer risk (Ross et al. 1998). In addition, several candidate genes for prostate cancer exhibit large allele frequency differences between African Americans and European Americans. These genes include the CAG-repeat tract within the androgen receptor gene (Irvine et al. 1995; Sartor et al. 1999; Kittles et al. 2001a), a TA-repeat tract within the *SRD5AR* gene (Reichardt et al. 1995), and the *CYP3A4* promoter variant examined here.

In this study, we have used a case-control association design to determine whether *CYP3A4-V* is associated with prostate cancer in African Americans after controlling for population stratification. First, we have examined whether the *CYP3A4* association observed among African Americans exists in other populations by comparing data from two populations ancestral to African Americans, Nigerians, and European Americans. Second, using the three groups of self-reported ethnicities, we performed Cochran-Mantel-Haenszel tests to determine if a possible association exists between the *CYP3A4* variant and prostate cancer. Then, to assess population stratification directly within the African American population, we have typed 10 unlinked (unrelated to prostate cancer) autosomal genetic markers, which, like the *CYP3A4* variant, exhibit large differences (>40%) in allele frequencies (δ) between Africans and Europeans (Parra et al. 1998, 2001). To correct for the stratification, we used the method of Reich and Goldstein (2001), which utilizes the association statistics observed at the 10 unlinked markers to determine whether the significant association at the candidate marker truly indicates the presence of a disease-related gene. We observed a strong association between the *CYP3A4* genotype and prostate cancer, in addition to significant population stratification in African Americans. The subsequent test for the association of *CYP3A4-V* with prostate cancer (controlling for the stratification) is not significant. Our results show that genetic association studies with African Americans are highly susceptible to confounding because of population stratification.

Subjects and methods

Study subjects

Unrelated men were enrolled from three sites for a population-based case-control study of risk factors for prostate cancer (Table 1). The study was approved by the Howard University Institutional Review Board and written consent was obtained from all subjects. All prostate cancer cases were between 40 to 85 years of age and were diagnosed with prostate cancer within the last 2 years. African Americans (n=220; 84 prostate cancer patients and 136 healthy

Table 1 Clinical characteristics of the prostate cancer (*Pca*) cases and healthy controls (*PSA* prostate-specific antigen, *NA* information not collected)

Population	*n*	Mean age (±SEM)	Mean PSA (±SEM)	Family history
African Americans	220			
Pca cases	84	67.8 (1.2)	24.3 (4.7)	15 (18%)
Controls	136	63.3 (0.8)	2.3 (0.6)	27 (20%)
Nigerians	159			
Pca cases	77	73.1 (1.1)	169.8 (19.1)	NA
Controls	82	71.9 (1.6)	1.5 (0.5)	NA
European Americans	309			
Pca cases	215	63.4 (0.7)	20.4 (2.7)	21 (10%)
Controls	94	60.1 (0.6)	1.2 (0.4)	6 (7%)

male controls) were recruited from the Washington, DC area through the Division of Urology at the Howard University Hospital and/or prostate cancer screening at the Howard University Cancer Center. The response rate among the African American cases was 92%. Healthy African American male volunteers were enrolled among individuals undergoing regular physical examinations at the Division of Urology at Howard University Hospital and/or men participating in screening programs for prostate cancer at the Howard University Cancer Center. The response rate for the African American controls was 90%. European Americans (n=309; 215 prostate cancer patients and 94 healthy male volunteers) were enrolled through various prostate-cancer-screening programs in Baltimore, sponsored by the Johns Hopkins Cancer Center. For the African Americans and European American participants, ethnicity was self-reported, and individuals of mixed ancestry were not excluded. In addition, Nigerians (n=159; 77 prostate cancer patients and 82 healthy male controls) belonging to the Yoruba ethnic group were enrolled through the University College Hospital in Ibadan, Nigeria and through a community-based study of risk factors for prostate cancer during the summer of 2000 in collaboration with the University of Benin Teaching Hospital in Benin City, Nigeria. The response rate among the Nigerian controls was 85%. Blood samples were collected from each subject. Clinical characteristics, including Gleason grade, prostate-specific antigen (PSA), TNM stage, age at diagnosis, and family history, were obtained from medical records. A positive family history was determined by having a first-degree relative affected with prostate cancer. Among the African American subjects, 18% of the prostate cancer cases and 20% of the healthy volunteers reported a positive family history of prostate cancer. For the European American subjects, 10% of the prostate cancer cases and 7% of the healthy volunteers reported family history. No family history data was collected for the Nigerian samples. All healthy controls had PSA levels less than 4.0 ng/ml and normal digital rectal examinations. The mean age at diagnosis among all prostate cancer patients was 64±1.1 years. The mean age among the controls was 63±1.6 years.

Genotyping

Genomic DNA was isolated from lymphocytes by standard proteinase K digestion, cell lysis, protein precipitation, and DNA precipitation. Genotyping of the *CYP3A4* A to G polymorphism was performed by using polymerase chain reaction (PCR) and restriction digestion. PCR amplification of the polymorphism was carried out with 200 nM forward primer (5'-GGA CAG CCA TAG AGA CAA GGG GA-3') and 200 nM reverse primer (5'-CAC TCA CTG ACC TCC TTT GAG TTC A-3'), which produced a 190-bp fragment. The PCR mix consisted of 30 ng genomic DNA, 12 pmol each primer, 1.25 U AmpliTaq polymerase (Perkin Elmer, Foster City, Calif.), 10× PCR buffer II (Perkin Elmer), 1.6 mM $MgCl_2$, 0.7 mM dNTP, and 10% dimethylsulfoxideDMSO (Sigma, St. Louis, Mo.) in a total volume of 25 µl. Reaction conditions included an initial melting step at 95°C for 5 min, followed by 35 cycles of melting at 95°C for 30 s, annealing at 60°C for 25 s, and extending at 72°C for 30 s. A final extension was set at 72°C for 4 min. Restriction enzyme digestion was performed on the PCR fragment in 10 µl PCR product, 2 µl 10× buffer II, 0.2 µl 100× bovine serum albumin, 2 µl (10 U) *Mbo*II (New England BioLabs), and 5.8 µl water in a total volume of 20 µl and incubated at 37°C overnight. The resultant fragments were electrophoresed on a 4% agarose gel containing ethidium bromide. Bands were then visualized by UV trans-illumination. All samples were assayed in duplicate directly from genomic DNA together with a set of control DNAs that included known homozygous AA and GG and heterozygous AG genotypes. These control DNAs were confirmed by direct DNA sequencing in an ABI 377 DNA sequencer (Applied Biosystems, Foster City, Calif.).

In addition, ten autosomal markers (*APOA1*, *AT3*, *FY*, *ICAM1*, *LPL*, *D11S429*, *OCA2*, *RB1*, *Sb19.3*, and *GC*) were genotyped in the African American samples by standard PCR and electrophoretic separation of DNA fragments. *APOA1* and *Sb19.3* are ALU polymorphisms, *AT3* is a 68-bp insertion/deletion polymorphism, and *FY*, *ICAM1*, *LPL*, *OCA2*, *RB1*, *GC*, and *D11S429* are SNPs typed by restriction enzymes. The primer sequences and PCR conditions for the ten loci are described in detail in Parra et al. (1998, 2001).

Statistical analysis

Genotype and allele frequencies were calculated for each population. Hardy-Weinberg equilibrium analyses for each population were evaluated by contingency table analysis. Two-sided Pearson chi-square (χ^2_p), odds ratios, and P-values were determined from comparisons of individual and combined genotype classes between prostate cancer patients and healthy controls for each of the three populations. Regression analyses were used to assess whether age at diagnosis and family history modified the relationship between *CYP3A4* and prostate cancer risk. In addition, the Cochran-Mantel-Haenszel χ^2 statistic was used to test for association of prostate cancer after adjusting for the different ethnic groups. To detect stratification within the African American population, Pearson χ^2 tests for association with prostate cancer were performed on genotypes at each of the 10 unlinked markers. The sum of the test statistics for each locus was then computed with the number of degrees of freedom (df) being equal to the sum of the number of df of the individual loci (Pritchard and Rosenberg 1999).

To correct for the stratification, the mean of the unlinked marker test statistics $\left(\bar{\chi}^2\right)$ was determined. A 95% upper confidence limit on the mean value (μ) was determined by multiplying the mean by 1.83 (based on using 10 markers and the χ^2 distribution in the absence of stratification; see Reich and Goldstein 2001). The candidate marker (*CYP3A4*) χ^2_p value was divided by the adjusted mean (μ) of the unlinked makers resulting in a χ^2 corrected for stratification (χ^2_{corr}) and a conservative P-value. Regression analyses were used to compare Gleason grade and TNM stage among prostate cancer patients. Prostate cancer patients were defined as "Low", i.e., a T1a-T1c and/or Gleason grade larger than 7, or "High", i.e., the T2–T4 or N (+) or M (+) stage and/or the Gleason grade equal to or larger than 7 (see Rebbeck et al. 1998; Paris et al. 1999). In the regression model, age at diagnosis, PSA (total), and family history (affected first degree relative) were controlled for among prostate cancer patients. The SAS Version 6.12 computer program (SAS Institute, Cary, N.C.) was used to compute all χ^2 tests, odds ratios, and P-values.

Results

CYP3A4-V frequencies across populations

The *CYP3A4-V* allele frequency was highest among Nigerians (87%), lowest among European Americans (10%), and intermediate among African Americans (66%). Previous reports (Paris et al. 1999; Walker et al. 1998) estimate the allele frequency for African Americans at about 53%. The higher *CYP3A4* allele frequency in our sample of unrelated African Americans may be attributable to differences in levels of admixture among geographically diverse African American communities. Genotype frequencies of *CYP3A4-V* in the three ethnic populations are shown in Table 2. Genotypes in each population were in Hardy-Weinberg equilibrium (P>0.05). Genotype frequencies differed significantly among the three control populations (P<0.001).

Table 2 CYP3A4 genotype frequencies in prostate cancer cases and healthy controls from three populations (*n* number of subjects, *OR* odds ratio, *CI* confidence interval). The two-sided *P* was obtained from Pearson χ^2 tests

Genotype	Cases	Controls	OR	95% CI	*P*-value
African Americans	*n*=84	*n*=136			
AA	4 (5%)	23 (17%)	1.0 (ref)		
AG	32 (38%)	44 (32%)	4.2	1.3–13.2	0.01
GG	48 (57%)	69 (51%)	4.0	1.3–12.3	0.01
AG+GG	80 (95%)	113 (83%)	4.1	1.3–12.2	0.007
European Americans	*n*=215	*n*=94			
AA	161 (75%)	82 (88%)	1.0 (ref)		
AG	28 (13%)	6 (6%)	2.3	0.9–5.9	0.06
GG	26 (12%)	6 (6%)	0.8	0.8–5.6	0.09
AG+GG	54 (25%)	12 (12%)	2.3	1.1–4.5	0.02
Nigerians		*n*=77	*n*=82		
AA	3 (4%)	1 (1%)	1.0 (ref)		
AG	23 (30%)	20 (24%)	0.4	0.3–3.9	0.41
GG	51 (66%)	61 (75%)	0.3	0.1–2.8	0.25
AG+GG	74 (96%)	81 (99%)	0.3	0.1–2.9	0.28

Uncorrected association of *CYP3A4-V* with prostate cancer

The relationship between *CYP3A4-V* and prostate cancer is presented in Table 2. Genotypes were compared among the three individual populations of cases and controls. For the African Americans, a strong association was observed with *CYP3A4* genotype. Individuals with at least one copy of the variant allele were at increased risk for prostate cancer, with an odds ratio (OR) of 4.1 and a 95% confidence interval (CI) of 1.3–12.2 (P=0.007). Further analyses revealed no relationship with *CYP3A4* and age of diagnosis or family history among the African Americans. An OR of 2.3 (95% CI 1.1–4.5; P=0.02) was observed for *CYP3A4* genotype and prostate cancer among European Americans; however, the *P*-value was not significant after correction for multiple-tests. Whereas the ORs were 2.3 (95% CI 0.9–5.9) and 0.8 (95% CI 0.8–5.6) for the AG and GG genotypes, respectively, among European Americans, they failed to be significant (P>0.06). Similar to African Americans, the age of diagnosis and family history among the European Americans were not associated with *CYP3A4* genotype (data not shown). No association between *CYP3A4* genotype and prostate cancer was observed among Nigerians (P=0.28). The variant allele was common among Nigerians with 96% (74 of 77) of prostate cancer patients and a striking 99% (81 of 82) of healthy controls possessing at least one copy of the allele (Table 2). Because of the high frequency of *CYP3A4-V*, a much larger sample size would be needed to detect an association with prostate cancer.

The standard Cochran Mantel Haenszel χ^2 test was employed to test for association between *CYP3A4-V* and prostate cancer in an attempt to control for differences among the three populations. Results of the analysis also indicated a strong association between the *CYP3A4* genotype and prostate cancer while controlling for differences among the three populations of self-reported ethnicities

Table 3 Comparison of CYP3A4 genotype with grade/stage[a] among prostate cancer patients (*n* number of subjects, *OR* odds ratio, *CI* confidence interval). The *P*-value was taken from logistic regression analyses controlling for age, PSA, and family history

Genotype	Low	High	OR	95% CI	*P*-value
African Americans	*n*=33	*n*=51			
AA	2 (6%)	2 (4%)	1.0 (ref)		
AG	15 (45%)	17 (33%)	0.9	0.1–7.0	0.90
GG	16 (49%)	32 (63%)	0.5	0.1–3.8	0.50
AG + GG	31 (94%)	49 (96%)	0.6	0.1–4.7	0.65
European Americans	*n*=129	*n*=82			
AA	90 (70%)	66 (80%)	1.0 (ref)		
AG	31 (24%)	13 (16%)	1.7	0.9–3.5	0.12
GG	8 (6%)	3 (4%)	1.9	0.4–7.6	0.33
AG+GG	39 (30%)	16 (20%)	1.8	0.9–3.4	0.08
Nigerians		*n*=8	*n*=69		
AA	1 (13%)	2 (3%)	1.0 (ref)		
AG	1 (13%)	22 (32%)	0.9	0.1–2.0	0.07
GG	6 (74%)	45 (65%)	0.3	0.2– 3.4	0.28
AG+GG	7 (87%)	67 (97%)	0.2	0.1–2.6	0.18

[a]Grade/stage as defined as *Low* (T1a-T1c stage and/or Gleason grade <7) or *High* (T2–T4 or N (+) or M (+) stage and Gleason grade ≥7)

(χ^2=10.07, P=0.002; OR= 2.35, 95% CI 1.4–3.9). Stratified analyses of *CYP3A4* genotypes and clinical characteristics in the ethnic populations are shown in Table 3. These analyses, which controlled for age, PSA, and family history, revealed that no association was observed between *CYP3A4* genotype and the combined Gleason grade and TNM stage in any of the three ethnic populations. Sample sizes within several of the categories, particularly among the Nigerians were low and may have contributed to our inability to detect a relationship between genotype and clinical characteristics.

Testing and correcting for population stratification

We tested for population stratification by comparing 10 unlinked autosomal genetic markers with prostate cancer in African Americans. Table 4 reveals that three of the 10 marker loci were nominally to strongly associated with prostate cancer in African Americans: *GC*1F* (P=0.003), *OCA2*1* (P=0.020), and *RB1*1* (P=0.047). *GC*1*F is one of several alleles at the group-specific component locus (Mastana et al. 1996), and *OCA2*1* is a SNP in exon 10 of the P-gene, a transporter protein involved in melanogenesis (Lee et al. 1995). *RB1*1* is a polymorphism within the tumor suppressor retinoblastoma gene (Zheng and Lee 2001). Given what is known about the function of these three genes, it is unlikely that any of them plays a role in prostate cancer. Overall, the test for population stratification using all 10 of the unlinked markers was highly significant (χ^2=29.9; df=10; P=0.008; Table 4).

The individual test statistics for the 10 unlinked markers were also used to correct for the population stratification in the African American samples. Specifically, the mean of the 10 marker χ^2_p statistics (2.98) was used to correct the initial Pearson χ^2 calculated for the *CYP3A4* comparison with prostate cancer in African Americans. This χ^2 statistic, which corrected for the level of stratification, was not significant (χ^2_{corr}=1.3, P=0.25).

Table 4 Test for stratification by comparison of unlinked markers[a] with prostate cancer in African Americans

Marker	Locus	χ^2	P-value[b]
APOA1*1	11q23	3.75	0.055
AT3*1	1q23–q25	2.40	0.122
GC*1F	4q12–q13	12.79	0.0003
FY-Null*1	1q22–q23	0.06	0.806
ICAM1*1	19p13	0.05	0.820
LPL*1	8p22	0.72	0.395
OCA2*1	15q11.2–q12	5.38	0.020
RB1*1	13q14.3	3.93	0.047
SB19.3*1	19	0.0002	0.987
D11S429*1	11	0.82	0.365
TOTAL		29.9	0.0008

[a]All loci are unlinked genetically, except FY and AT3, which are ~22 cM apart on chromosome 1
[b]Two-sided P-value from Pearson χ^2 tests

Discussion

Even though case-control association studies may be powerful for detecting the non-random association between an allele and a trait, they are vulnerable to confounding because of population stratification (Chakraborty and Weiss 1988; Lander and Schork 1994). Population stratification can be caused by various circumstances (Reich and Goldstein 2002). One example is when two or more groups with different allele frequencies are pooled together in the case and control samples under study. In addition, admixture can create population stratification. This is especially the case for the African American population, which, because of its unique population history, represents a varied mixture of African, European, and Native American ancestry. The problem of stratification is compounded when the disease of interest is more prevalent in one of the populations, as is the case with prostate cancer among African Americans. Any alleles that are more common among African Americans will tend to be associated with the disease, even if it is completely unlinked to the disease-causing locus. Several approaches have been attempted that deal with the problem of population stratification. One approach is to match the ethnic backgrounds of cases and controls. However, considerable "cryptic" or hidden stratification may still remain (Ewens and Spielman 1995). Another approach has been to collect controls from families of affected individuals. Depending on the study, family-based control methods such as the transmission disequilibrium test (TDT) and a related method (sib-TDT; Ewens and Spielman 1995; Spielman and Ewens 1998) are more difficult and costly than collecting unrelated cases and controls. For diseases with a late age of onset, such as prostate cancer, the availability of parents and siblings for sampling is greatly reduced. Furthermore, in some instances, TDT-type designs may exhibit less power compared with case-control studies because of over-matching of unaffected sibs to probands (Risch 2000; Risch and Teng 1998; Morton and Collins 1998). Recently, it has been proposed that unlinked genetic markers should be typed in order to detect, quantify, and correct for stratification in the case-control study (Pritchard and Rosenberg 1999; Pritchard et al. 2000; Devlin and Roeder 1999; Schork et al. 2001; Reich and Goldstein 2001).

In this study, we have shown that population stratification is a potential problem for association studies in the African American population when there are differences in allele frequencies between the parental populations. Our results on *CYP3A4-V*, a candidate gene polymorphism for prostate cancer, have revealed that the promoter allele frequency differs significantly between populations ancestral to African Americans. Not surprisingly, a strong association was observed between *CYP3A4* genotype and prostate cancer in African Americans. This association is consistent with previous studies of *CYP3A4* and prostate cancer in African Americans (Paris et al. 1999). In the previous work, population substructure within the African

American population was not evaluated even though the *CYP3A4-V* frequency varied substantially among the different African American populations under study (Ball et al. 1999; Walker et al. 1998; Paris et al. 1999). In this study, we have provided evidence for population stratification within the African American population from Washington, DC, by performing tests of association with 10 unlinked autosomal genetic markers. These analyses revealed that three of the 10 marker alleles are also significantly associated with prostate cancer. Reich and Goldstein (2001) suggest that the unlinked markers should be matched to the candidate locus based on allele frequency. However, we have used a panel of 10 unlinked genetic markers similar to the candidate gene locus in terms of large δ-values (>40%) between Africans and Europeans. The panel of unlinked genetic markers provides a more powerful tool for detecting genetic substructure (stratification) within the African American population than randomly chosen markers with allele frequencies similar to the candidate locus (Pfaff et al. 2002; Reich and Goldstein 2002). The substructure detected is attributable to differences in ancestral proportions between the prostate cancer cases and clinically evaluated controls. It is likely that stratification in the African American samples may have resulted in a spurious association of *CYP3A4-V* with prostate cancer in our samples. If so, the confounding would have to be strong because, even after taking into account self-reported ethnicity by using the Cochran-Mantel-Haenszel χ^2 analysis, the association with the *CYP3A4* genotype in the African American population is still quite significant. This may reflect the inherit problem of "cryptic" stratification when using self-reported "race" or ethnicity in grouping individuals for genetic epidemiological studies.

This is the first study that, while examining the role that a genetic variant plays in the etiology of prostate cancer in African Americans, also assesses and corrects for population stratification within the population. In order to determine the roles that genes such as *CYP3A4-V* play in the etiology of prostate cancer among African Americans, methods that deal with the issue of stratification are important because of differences in allele frequencies and disease prevalence among populations. In terms of *CYP3A4-V* specifically, more research into the functional consequence of the *CYP3A4* promoter variant is needed. It is possible that the polymorphism leads to an altered form of the transcriptional regulatory element (NSFE) possibly affecting gene expression that ultimately may results in the decreased oxidation of testosterone. Although some research supports altered function, there is no consensus (Westlind et al. 1999; Amirimani et al. 1999; Ando et al. 1999; Ball et al. 1999; Rebbeck 2000), and we have not found strong supportive evidence in our association studies of the European American and Nigerian clinical samples.

The results for AG and GG genotype among European Americans were suggestive before correction for multiple tests (OR=2.3; P=0.02). The trend was intriguing and warranted a test for stratification in the European American samples but, because of limited genomic DNA, we were unable to genotype all of the European American samples. We were however able to type the unlinked markers in individuals with prostate cancer who possessed at least one copy of the *CYP3A4-V* allele (n=54). Surprisingly, the results suggested that 15% of these individuals also possessed at least one copy of the African-associated alleles at the *FY* and *GC* marker loci. This being the case, these individuals probably have considerable African ancestry but classify themselves as European Americans.

These interesting results need to be confirmed by typing a larger number of unlinked markers with high δs. As many of these markers are widely becoming available, they will improve the power to detect and control for population stratification (Parra et al. 1998, 2001; Pfaff et al. 2001, 2002; Pritchard and Rosenberg 1999). In addition, these markers can be utilized for genetic approaches, such as mapping by admixture linkage disequilibrium, a potentially useful application for the identification of genes contributing to complex genetic diseases (Briscoe et al. 1994; Stephens et al. 1994; Smith et al. 1996; Zheng and Elston 1999; Pfaff et al. 2001, 2002; Collins-Schramm et al. 2002). Increasing the number of unlinked genetic markers that are informative for ancestry should also allow us to estimate individual admixture in the African American population. An interesting and potentially powerful utilization of the individual admixture estimates would be as independent risk factors on which cases and controls can be matched for analysis (Williams et al. 2000). Subsequently, stratification would be minimized, and relationships between prostate cancer and candidate loci would be better evaluated. We intend to explore this relationship in a larger dataset in the future.

Our inability to detect an association among the Nigerian samples may be attributable to several reasons. The first is the high frequency of the variant in the population. The high frequency of the *CYP3A4-V* allele among Nigerians possibly contributed to the difficulty in observing possible low penetrant effects in the Nigerian population. A much larger sample size may be needed in order to detect a genetic effect. Another explanation may be that specific genotype-environment interactions are absent or of low effect in rural Nigeria because of diet and lifestyle differences from those of African Americans. Even though population stratification may have contributed to the association with prostate cancer in the African American samples, these scenarios cannot be ruled out as likely explanations for the lack of association within the Nigerian and European American samples.

Future studies may provide greater knowledge of the role, if any, that the *CYP3A4* gene plays in prostate cancer development and progression. However, other genes or genomic regions have been identified that may contribute to the susceptibility to prostate cancer, such as the highly penetrant but less common alleles at the *HPC1* (Smith et al. 1996), *HPC2* (Tavtigan et al. 2001) and *HPCX* (Xu et al. 1998) loci and the common low penetrant alleles at the androgen receptor (Giovannucci et al. 1997) *SRD5AR* (Jaffe et al. 2000) and *CYP17* (Habuchi et al. 2000; Gsur et al. 2000; Kittles et al. 2001b) genes.

Acknowledgements We are grateful to Dr. U. Osime, N. Foster, N. Sylvester, and D. Young for their assistance in subject enrollment and data processing. We also thank Drs. G. Bonney and C. Rotimi for helpful discussions and support with statistical methods and Drs. M. Shriver and E. Parra for providing input on genetic markers. This work was supported by grants RR03048-13S1 from NCRR, National Institutes of Health and DAMD17-00-1-0025 from the Department of Defense, and by the Howard University Cancer Center.

References

Amirimani B, Walker AH, Weber BL, Rebbeck TR (1999) Response to Ando et al. Re: Modification of clinical presentation of prostate tumors by a novel genetic variant in CYP3A4. J Natl Cancer Inst 91:1588–1590

Ando Y, Tateishi T, Sekido Y, Yamamoto T, Satoh T, Hasegawa Y, et al (1999) Re: Modification of clinical presentation of prostate tumors by a novel genetic variant in CYP3A4. J Natl Cancer Inst 91:1587–1588

Ball SE, Scatina J, Kao J, Ferron GM, Fruncillo R, Mayer P, et al (1999) Population distribution and effects on drug metabolism of a genetic variant in the 5' promoter region of CYP3A4. Clin Pharmacol Ther 66:288–294

Briscoe D, Stephens JC, O'Brien SJ (1994) Linkage disequilibrium in admixed populations: applications in gene mapping. J Hered 85:59–63

Chakraborty R, Weiss KM (1988) Admixture as a tool for finding linked genes and detecting that difference from allelic association between loci. Proc Natl Acad Sci USA 85:9119–9123

Collins FS, Guyer MS, Chakravarti A (1997) Variations on a theme: cataloging human DNA sequence variation. Science 278: 1580–1581

Collins-Schramm H, Phillips C, Operario D, Lee J, Weber J, Hanson R, Knowler W, Cooper R, Li H, Seldin M (2002) Ethnic-difference markers for use in mapping by admixture linkage disequilibrium. Am J Hum Genet 70:737–750

Devlin B, Roeder K (1999) Genomic control for association studies. Biometrics 55:997–1004

Ewens WJ, Spielman RS (1995) The transmission/disequilibrium test: history, subdivision, and admixture. Am J Hum Genet 57:455–464

Giovannucci E, Stampfer MJ, Krithivas K, Brown M, Dahl D, Brufsky A, Talcott J, Hennekens CH, Kantoff PW (1997) The CAG repeat within the androgen receptor gene and its relationship to prostate cancer. Proc Natl Acad Sci USA 94:3320–3323

Greenlee RT, Murray T, Bolden S, Wingo PA (2000) Cancer statistics, 2000. Cancer J Clin 50:7–33

Gsur A, Bernhofer G, Hinteregger S, Haidinger G, Schatzl G, Madersbacher S, Marberger M, Vutuc C, Micksche M (2000) A polymorphism in the CYP17 gene is associated with prostate cancer risk. Int J Cancer 87:434–437

Habuchi T, Liqing Z, Suzuki T, Sasaki R, Tsuchiya N, Tachiki H, Shimoda N, Satoh S, Sato K, Kakehi Y, Kamoto T, Ogawa O, Kato T (2000) Increased risk of prostate cancer and benign prostatic hyperplasia associated with a CYP17 gene polymorphism with a gene dosage effect. Cancer Res 60:5710–5713

Irvine RA, Yu MC, Ross RK, Coetze GA (1995) The CAG and GGC microsatellites are in linkage disequilibrium in men with prostate cancer. Cancer Res 55:1937–1940

Jaffe J, Malkowicz B, Walker A, MacBride S, Peschel R, Tomaszewski J, Arsdalen K, Wein A, Rebbeck T (2000) Association of SRD5A2 Genotype and Pathological Characteristics of Prostate Tumors. Cancer Research 60:1626–30

Kittles R, Young D, Weinrich S, Hudson J, Aryropoulos G, Ukoli F, Adams-Campbell L, Dunston G (2001a) Extent of linkage disequilibrium between the androgen receptor gene CAG and GGC repeats in human populations: implications for prostate cancer risk. Hum Genet 109:253–261

Kittles R, Panguluri R, Chen W, Massac A, Ahaghotu C, Jackson A, Ukoli F, Adams-Campbell L, Isaacs W, Dunston G (2001b) CYP17 promoter variant associated with prostate cancer aggressiveness in African Americans. Cancer Epidemiol Biomarkers Prev 10:943–947

Kruglyak L (1999) Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nat Genet 22: 139–144

Lander ES, Schork NJ. (1994) Genetic dissection of complex traits. Science 265:2037–2048

Lee S-T, Nicholls R, Jong M, Fukai K, Spritz R (1995) Organization and sequence of the human P gene and identification of a new family of transport proteins. Genomics 26:354–363

Mastana SS, Constans J, Papiha SS (1996) New rare variants and interpretation of the genetic diversity at the group-specific component GC locus in populations of India. Hum Biol 685: 795–818

Morton NE, Collins A (1998) Tests and estimates of allelic association in complex inheritance. Proc Natl Acad Sci USA 95: 11389–11393

Paris PL, Kupelian PA, Hall JM, Williams TL, Levin H, Klein EA, Casey G, Witte JS (1999) Association between a CYP3A4 genetic variant and clinical presentation in African-American prostate cancer patients. Cancer Epidemiol Biomarkers Prev 8:901–905

Parra EJ, Marcini A, Akey J, Martinson J, Batzer MA, Cooper R, Forrester T, Allison DB, Deka R, Ferrell RE, Shriver MD (1998) Estimating African admixture proportions by use of population-specific alleles. Am J Hum Genet 63:1839–1851

Parra EJ, Kittles R, Argyropoulos G, Hiester K, Bonilla C, Sylvester N, Parrish-Gause D, Garvey W, Kamboh M, Jin L, Ferrell R, Pollitzer W, Mckeighe P, Shriver MD (2001) Ancestral proportions and admixture dynamics in geographically defined African Americans living in South Carolina. American Journal of Physical Anthropology 114:18–29

Pfaff C, Parra E, Bonilla C, Hiester K, McKeigue P, Kamboh M, Hutchinson R, Ferrell R, Boerwinkle E, Shriver MD (2001) Population structure in admixed populations: effect of admixture dynamics on the pattern of linkage disequilibrium. Am J Hum Genet 68:198–207

Pfaff C, Kittles R, Shriver MD (2002) Adjusting for population structure in admixed populations. Genet Epidemiol 22:196–198

Pritchard JK, Rosenberg NA (1999) Use of unlinked genetic markers to detect population stratification in association studies. Am J Hum Genet 65:220–228

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using mutlilocus genotype data. Genetics 155: 945–959

Rebbeck TR (2000) More about: modification of clinical presentation of prostate tumors by a novel genetic variant in CYP3A4. J Natl Cancer Inst 92:76

Rebbeck TR, Jaffe JM, Walker AH, Wein AJ, Malkowicz SB (1998) Modification of clinical presentation of prostate tumors by a novel genetic variant in CYP3A4. J Natl Cancer Inst 90: 1225–1229

Reich D, Goldstein D (2001) Detecting association in a case-control study while correcting for population stratification. Genet Epidemiol 20:4–16

Reich D, Goldstein D (2002) Re: Adjusting for population structure in admixed populations. Genet Epidemiol 22:198–200

Reichardt JKV, Makridakis N, Henderson B, Yu M, Pike M, Ross R (1995) Genetic variability of the human SRD5A2 gene: implications for prostate cancer risk. Cancer Res 55:3973–3975

Ries LAG, Eisner MP, Kosary CL, Hankey BF, Miller BA, Edwards BK (2000) SEER Cancer Statistics Review, 1973–1997. National Cancer Institute, Bethesda, Md., USA

Risch NJ (2000) Searching for genetic determinants in the new millennium. Nature 405:847–856

Risch NJ, Teng J (1998) The relative power of family-based and case-control designs for linkage disequilibrium studies of complex human diseases. I. DNA pooling. Genome Res 8:1273–1288

Ross RK, Coetzee GA, Reichardt JK, Skinner E, Henderson BE (1995) Does the racial-ethnic variation in prostate cancer risk have a hormonal basis? Cancer 75:1778–1782

Ross RK, Pike MC, Coetzee GA, Reichardt JK, Yu MC, Feigelson H, Stanczyk FZ, Kolonel LN, Henderson BE (1998) Androgen metabolism and prostate cancer: establishing a model of genetic susceptibility. Cancer Res 58:4497–4404

Sartor O, Zheng Q, Eastham J (1999) Androgen receptor gene CAG repeat length varies in a race-specific fashion in men without prostate cancer. Urology 53:378–380

Schork NJ, Fallin D, Thiel B, Xu X, Broeckel U, Jacob HJ, Cohen D (2001) The future of genetic case-control studies. Adv Genet 42:191–212

Smith JR, Freije D, Carpten JD, Xu J, Isaacs SD, Brownstein MJ, et al (1996) Major susceptibility locus for prostate cancer on chromosome 1 suggested by a genome-wide search. Science 274:1371–1374

Spielman RS, Ewens WJ (1998) A sibship test for linkage in the presence of association: the sib transmission/disequilibrium test. Am J Hum Genet 62:450–458

Stephens JC, Briscoe D, O'Brien SJ (1994) Mapping by admixture linkage disequilibrium in human populations: limits and guidelines. Am J Hum Genet 55:809–824

Tavtigian SV, Simard J, Teng DH, Abtin V, Baumgard M, Beck A, Camp NJ, Carillo AR, Chen Y, Dayananth P, et al (2001) A candidate prostate cancer susceptibility gene at chromosome 17p. Nat Genet 27:172–180

Wacholder S, Rothman N, Caporaso N (2000) Population stratification in epidemiologic studies of common genetic variants and cancer: quantification of bias. J Natl Cancer Inst 92:1151–1158

Walker AH, Jaffe JM, Gunasegaram S, Cummings SA, Huang C, Chern H, Olopade OI, Weber BL, Rebbeck TR (1998) Characterization of an allelic variant in the nifedipine-specific element of CYP3A4: ethnic distribution and implications for prostate cancer risk. Hum Mutat 12:289

Waxman DJ, Attisano C, Guengerich FP, Laperson DP (1998) Human liver microsomal steriod metabolism: identification of the major microsomal steriod hormone 6β-hydroxylase cytochrome P-450 enzyme. Arch Biochem Biophys 263:424–436

Westlind A, Loftberg L, Tindberg N, Anderson TB, Ingelman-Sundberg M (1999) Interindividual differences in hepatic expression of CYP3A4: relationship to genetic polymorphism in the 5'-upstream regulatory region. Biochem Biophys Res Commun 259:201–205

Williams RC, Long JC, Hanson RL, Sievers ML, Knowler W (2000) Individual estimates of European genetic admixture associated with lower body-mass index, plasma glucose, and prevalence of type 2 diabetes in Pima Indians. Am J Hum Genet 66:527–538

Xu J, Meyers D, Freije D, Isaacs S, Wiley S, Nusskern D, Ewing C, et al (1998) Evidence for a prostate cancer susceptibility locus on the X chromosome. Nat Genet 20:175–179

Zheng C, Elston R (1999) Multipoint linkage disequilibrium mapping with particular reference to the African American population. Genet Epidemiol 17:79–101

Zheng L, Lee WH (2001) The retinoblastoma gene: a prototypic and multifunctional tumor suppressor. Exp Cell Res 264:2–18